

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:35:16 ; Search time 13.43 Seconds
(without alignments)
960.242 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MRLGLLQTLRFTSPPTD.....CARGRVETNNEGPVSVMF 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107.5	5.9	1533	1 PUM_DROME	P25822 drosophila
2	105	5.8	480	1 TGN5_HUMAN	Q43493 homo sapien
3	102	5.6	1914	1 KML5_HUMAN	Q15746 homo sapien
4	97	5.3	980	1 BOB1_YEAST	P38041 saccharomyc
5	96	5.3	367	1 CSP_PLAYO	P06914 plasmodium
6	96	5.3	892	1 ATX7_HUMAN	Q15265 homo sapien
7	95	5.2	1168	1 MYSC_ACACA	P10569 acanthamoeb
8	94.5	5.2	509	1 SOX9_PIG	Q18896 sus scrofa
9	94.5	5.2	602	1 CRK_DAUCA	P33681 daucus caro
10	93	5.1	507	1 MEFA_HUMAN	Q02078 homo sapien
11	93	5.1	509	1 SOX9_HUMAN	P48436 homo sapien
12	93	5.1	579	1 SYN3_RAT	O70441 rattus norv
13	93	5.1	604	1 RCOL1_NEUCR	P78706 neurospora
14	93	5.1	1329	1 FTSK_ECOLI	P46889 escherichia
15	92.5	5.1	498	1 MEFA_MOUSE	Q60929 mus musculu
16	92	5.1	363	1 TOBI_MOUSE	Q61471 mus musculu
17	91.5	5.0	411	1 Y513_HUMAN	O60268 homo sapien
18	91.5	5.0	785	1 SOK2_YEAST	P53438 saccharomyc
19	90.5	5.0	877	1 WEE1_SCHPO	P07527 schizosacch
20	90.5	5.0	1047	1 HIRA_DROME	O17468 drosophila
21	90.5	5.0	2515	1 TUD_DROME	P25823 drosophila
22	89.5	4.9	1460	1 PMPC_CHLMU	Q9pjvl chlamydia m
23	89	4.9	673	1 ESR2_MICUN	P57781 micropogoni
24	88.5	4.9	393	1 SMA3_CAEEL	P45896 caenorhabdi
25	88.5	4.9	649	1 INVA_PHAAR	P29001 phaeocelus a
26	88.5	4.9	1262	1 CAL3_CHICK	P12105 gallus gall
27	88	4.8	345	1 TOBI_HUMAN	P50616 homo sapien
28	88	4.8	734	1 YH09_YEAST	P32900 saccharomyc
29	88	4.8	838	1 GLT4_WHEAT	P08489 triticum ae
30	87.5	4.8	580	1 SYN3_HUMAN	O14994 homo sapien
31	87.5	4.8	1527	1 CALH_MOUSE	P39061 mus musculu
32	87	4.8	304	1 UL49_HSVEB	P28960 equine herp
33	87	4.8	609	1 RFA1_XENLA	Q01588 xenopus lae

RESULT 1

ID	PUM_DROME	STANDARD;	PRT;	1533 AA.
AC	P25822;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Maternal pumilio protein.			
GN	PUM.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=92249205; Pubmed=1576962;			
RA	Macdonald P.M.;			
RT	*The Drosophila pumilio gene: an unusually long transcription unit			
RT	and an unusual protein.*;			
RL	Development 114:221-234(1992).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93093466; Pubmed=1459455;			
RA	Barker D.D., Wang C., Moore J., Dickinson L.K., Lehmann R.;			
RT	*Pumilio is essential for function but not for distribution of the			
RT	Drosophila abdominal determinant Nanos.*;			
RL	Genes Dev. 6:2312-2326(1992).			
CC	-!- FUNCTION: PUM IS THE ONLY GENE REQUIRED FOR NOS ACTIVITY THAT IS			
CC	NOT ALSO REQUIRED FOR POSTERIOR LOCALIZATION OF GERM LINE			
CC	DETERMINANTS. PUM IS REQUIRED DURING EMBRYOGENESIS WHEN NOS			
CC	ACTIVITY APPARENTLY MOVES ANTERIORLY FROM THE POSTERIOR POLE.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC. IT IS CONCENTRATED IN THE			
CC	CORTICAL REGION OF THE EMBRYO BENEATH THE NUCLEI.			
CC	-!- DEVELOPMENTAL STAGE: EXPRESSED IN THE OVARIES AND DURING THE			
CC	EMBRYOGENESIS.			
CC	-!- DOMAIN: CONSISTS MAINLY OF REGIONS ENRICHED IN A SINGLE AMINO			
CC	ACID.			
CC	-!- DISEASE: LETHAL DEFECTIVE IN POSTERIOR PATTERN FORMATION.			
CC	-!- SIMILARITY: BELONGS TO THE PUMILIO/MPT5 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X62589; CAA44474.1; -			
DR	EMBL; L07943; AAB59189.1; -			
DR	PIR; S22026; S22026.			
DR	HSSP; P04002; IWPA.			
DR	Flybase; FBgn0003165; pum.			
DR	InterPro; IPR001313; PUM.			
DR	Pfam; PF00806; PUF; 8.			

PI4378 bovine herp
O24509 phaseolus v
PI1675 pseudorabie
O75410 homo sapien
Q9nyv4 homo sapien
P19056 rhodobacter
Q64127 mus musculu
P39060 homo sapien
Q08397 homo sapien
Q08r34 homo sapien
Q9nqc3 homo sapien
O14514 homo sapien

ALIGNMENTS

34 86.5 4.8 521 1 VGLC_HSVBC
35 86.5 4.8 651 1 INVA_PHAVU
36 86.5 4.8 1461 1 IE18_PRIVF
37 86 4.7 805 1 TAC1_HUMAN
38 86 4.7 1490 1 CRK7_HUMAN
39 85.5 4.7 254 1 RCEH_RHOCA
40 85.5 4.7 1051 1 TFIA_MOUSE
41 85.5 4.7 1516 1 CALH_HUMAN
42 85 4.7 574 1 LOLL_HUMAN
43 85 4.7 630 1 MIA3_HUMAN
44 85 4.7 1192 1 RTN4_HUMAN
45 84.5 4.6 1584 1 BAIL_HUMAN

CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
 CC THE DOMINANT SPICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
 CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
 CC -1- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U48959; AAC18423.2; -
 CC EMBL; AF069601; AAD15921.2; -
 CC EMBL; AF069602; AAD15922.1; -
 CC EMBL; AF069603; AAD15923.1; -
 CC EMBL; AF069604; AAD15924.1; -
 CC EMBL; X85337; CAA59685.1; -
 CC EMBL; AF096771; AAD51380.1; -
 CC EMBL; AF096766; AAD51380.1; JOINED.
 CC EMBL; AF096767; AAD51380.1; JOINED.
 CC EMBL; AF096768; AAD51380.1; JOINED.
 CC EMBL; AF096769; AAD51380.1; JOINED.
 CC EMBL; AF096770; AAD51380.1; JOINED.
 CC EMBL; AF096771; AAD51380.1; JOINED.
 CC EMBL; AF096774; AAD54018.1; -
 CC EMBL; AF096771; AAD51381.1; -
 CC EMBL; AF096770; AAD51381.1; JOINED.
 CC EMBL; AF096770; AAD51381.1; JOINED.
 CC HSP; P56276; 1TLK.
 CC MIM; 600922; -
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR003962; FNIII_repeat.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR003598; Ig_c2.
 CC InterPro; IPR003600; Ig_like.
 CC InterPro; IPR002290; Ser_thr_pkinase.
 CC Pfam; PF00041; fn3; 1.
 CC Pfam; PF00047; Ig; 8.
 CC Pfam; PF00069; pkinase; 1.
 CC PRINTS; PR00014; FNTYPEIII.
 CC SMART; SM00060; FN3; 1.
 CC SMART; SM00410; Ig_like; 1.
 CC SMART; SM00408; IGc2; 8.
 CC SMART; SM00220; S_TKG; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
 CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 CC Alternative initiation; Alternative splicing.
 CC CHAIN 1 1914
 CC MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
 CC ISOZYME.
 CC FT CHAIN 923 1914
 CC MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
 CC ISOZYME.
 CC FT CHAIN 1761 1914
 CC TELOKIN.
 CC FT INIT_MET 923 923
 CC FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-
 CC MUSCLE ISOZYME.
 CC FT INIT_MET 1761 1761
 CC FOR TELOKIN.
 CC FT DOMAIN 1343 1413
 CC FIBRONECTIN TYPE-III.
 CC FT DOMAIN 1464 1719
 CC PROTEIN KINASE.
 CC FT DOMAIN 1711 1774
 CC CALMODULIN-BINDING.
 CC FT DOMAIN 1824 1891
 CC IG-LIKE C2-TYPE DOMAIN.
 CC FT NP_BIND 1470 1478
 CC ATP (BY SIMILARITY).
 CC FT BINDING 1493 1493
 CC ATP (BY SIMILARITY).
 CC FT ACT_SITE 1585 1585
 CC BY SIMILARITY.
 CC FT DOMAIN 1906 1914
 CC POLY-GLU.
 CC FT DOMAIN 868 998
 CC 5 X 28 AA APPROXIMATE TANDEM REPEATS.

FT REPEAT 868 895 I-1.
 FT REPEAT 896 923 I-2.
 FT REPEAT 924 951 I-3.
 FT REPEAT 952 979 I-4.
 FT REPEAT 980 998 I-5 (INCOMPLETE).
 FT DOMAIN 999 1063 6 X 12 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 999 1003 II-1 (INCOMPLETE).
 FT REPEAT 1004 1015 II-2.
 FT REPEAT 1016 1027 II-3.
 FT REPEAT 1028 1039 II-4.
 FT REPEAT 1040 1051 II-5.
 FT REPEAT 1052 1063 II-6.
 FT VARSPLIC 437 506 VSGIPKPEVAMPLEGTQVRRQSGSTIEVYEDAGSHYLCILKA
 FT VARSPLIC 1433 1439 RTRDSGTCTASNAQGVSGSWTLQVER -> G (IN
 FT VARSPLIC 1473 1546 ISOFORM 2 AND ISOFORM 3B).
 FT VARSPLIC 1655 1705 DEVEUSD -> MKWRCT (IN ISOFORM 3A,
 FT VARSPLIC 1790 1790 ISOFORM 3B AND ISOFORM 4).
 FT CONFLICT 933 933 NCLHHPKLVQCVDAFEKAKANIVMWLEIVSGGEL -> L
 FT CONFLICT 963 963 (IN ISOFORM 4).
 FT CONFLICT 1022 1022 MISSING (IN ISOFORM 3A AND ISOFORM 3B).
 FT CONFLICT 1048 1050 V -> M (IN REF. 5).
 FT CONFLICT 1162 1162 S -> P (IN REF. 3; AAD15922).
 FT CONFLICT 1210 1210 P -> A (IN REF. 5).
 FT CONFLICT 1280 1280 KPM -> EAH (IN REF. 5).
 FT CONFLICT 1284 1284 P -> L (IN REF. 3; AAD15922 AND
 FT CONFLICT 1300 1300 AAD15923).
 FT CONFLICT 1316 1316 L -> P (IN REF. 5).
 FT CONFLICT 1326 1326 E -> D (IN REF. 3; AAD15922 AND
 FT CONFLICT 1478 1478 AAD15923).
 FT CONFLICT 1511 1511 M -> I (IN REF. 3; AAD15922 AND
 FT CONFLICT 1563 1563 AAD15924).
 FT CONFLICT 1609 1609 A -> G (IN REF. 5).
 FT CONFLICT 1639 1639 L -> S (IN REF. 5).
 FT CONFLICT 1639 1640 T -> S (IN REF. 5).
 FT CONFLICT 1639 1640 V -> C (IN REF. 5).
 FT CONFLICT 1639 1640 S -> T (IN REF. 3; AAD15922 AND
 FT CONFLICT 1639 1640 AAD15923).
 FT CONFLICT 1639 1640 I -> T (IN REF. 5).
 FT CONFLICT 1639 1640 A -> P (IN REF. 5).
 FT CONFLICT 1639 1640 G -> R (IN REF. 5).
 FT CONFLICT 1639 1640 GY -> D (IN REF. 3; AAD15922, AAD15923
 Query Match 5.6%; Score 102; DB 1; Length 1914;
 Best Local Similarity 22.4%; Pred. No. 9.1;
 Matches 58; Conservative 37; Mismatches 88; Indels 76; Gaps 14;
 QY 76 PLPAQ--PPALAOPOYOS-----POOPOTRWVAPRNNAFCQSGAGSDSNPGWOP 128
 DB 966 PVPEKVPVPKATPDPRSVLGGKKLPAENGSSAETLNAKAVESKPLSNAPSGPLKP 1025
 QY 129 --NSAPSVESHPV---LEKLKAHSYNPKFEENLKSQVFIKSYSEDDIHSIKYS 181
 DB 1026 VGNAPKPAETLKPNGNAKPAETLKPNGNAKPADE---NLKSA-----SKLELKDKVND 1074
 QY 182 IWCSTEH-----GNKRLDS-----AFR-----C--MSSKGPVYLLFSVN 213
 DB 1075 VNCKRGHACTTDNEXRSESGQTAPAFKQKLDVHVHAEKKLLLOQCVSSDPPATIIWTLN 1134
 QY 214 G-----SCHFCGVA-EMKSPVDYGTSGACVWSQDKWKKGFQVMTFVKDVPNN 259
 DB 1135 GKTLTKTKFIILSQEGSLGCVSIEKALPEDRGLYKCAVKNADAGACSCQ-VTVDDAPAS 1193
 QY 260 QLRHRLNNDNKPVTNSR 278
 DB 1194 E-----NTRAPENKSR 1204
 RESULT 4
 BOBL_YEAST
 ID BOBL_YEAST STANDARD: PRT; 980 AA.
 AC P38041;

FT DOMAIN 139 228 15 X 6 AA TANDEM REPEATS OF Q-G-P-G-A-P.
FT DOMAIN 229 260 8 X 4 AA TANDEM REPEATS OF Q-Q-P-P.
SQ SEQUENCE 367 AA: 38888 MW; 1EA56AFF7FECB5E3 CRC64;

Query Match 5.3%; Score 96; DB 1; Length 367;
Best Local Similarity 38.1%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 31; Indels 12; Gaps 5;

QY 72 PRAOPLPAQPPALAPQYQSPQPPQTRWVAPRNNAFGSCGAGSDSNSPCNVQPNRS- 130
Db 231 PQQBP-PQPP--QQPPQPPQPPQPPQPPR-----PPDGNNNNNNNGNNDNSY 281
QY 131 APSVESHPVLEKKAHSYNPKF 154
Db 282 VPSAEQ--ILEFVKQISSOLTEEM 303

RESULT 6
ATX7_HUMAN
ID ATX7_HUMAN STANDARD; PRT; 892 AA.
AC O15265; O75328; Q9Y6P8;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin 7 (Spinocerebellar ataxia type 7 protein).
GN SCA7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RX MEDLINE=97434213; PubMed=9288099;
RA David G., Abbas N., Stevanin G., Duerr A., Yvert G., Cancel G.,
RA Weber C., Imbert G., Saudou F., Antoniou E., Drabkin H., Gemmill R.,
RA Giunti P., Benomar A., Wood N., Ruberg M., Agid Y., Mandel J.-L.,
RA Brice A.;
RT "Cloning of the SCA7 gene reveals a highly unstable CAG repeat
RT expansion.";
RL Nat. Genet. 17:65-70(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=98087568; PubMed=9425224;
RA Del-Favero J., Krols L., Michalik A., Theuns J., Loeferen A.,
RA Goossens D., Wehnert A., van den Bossche D., van Zand K.,
RA Backhovens H., van Regenmortel N., Martin J.-J., Van Broeckhoven C.;
RT "Molecular genetic analysis of autosomal dominant cerebellar ataxia
RT with retinal degeneration (ADCA type II) caused by CAG triplet repeat
RT expansion.";
RL Hum. Mol. Genet. 7:177-186(1998).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99371772; PubMed=10441328;
RA Kaytor M.D., Duvick L.A., Skinner P.J., Koob M.D., Ranum L.P.,
RA Orr H.T.;
RT "Nuclear localization of the spinocerebellar ataxia type 7 protein,
RT ataxin-7.";
RL Hum. Mol. Genet. 8:1657-1664(1999).
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN ADDITION TO A DIFFUSE
CC DISTRIBUTION THROUGHOUT THE NUCLEUS, IT IS ASSOCIATED WITH THE
CC NUCLEAR MATRIX AND THE NUCLEOLUS.
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF SCA7 IS HIGHLY POLYMORPHIC (4
CC TO 18 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT
CC 30-130 REPEATS IN SCA7 PATIENTS. INTERMEDIATE ALLELES WITH 28 TO
CC 35 REPEATS ARE PRONE TO FURTHER EXPANSION.
CC -!- DISEASE: DEFECTS IN SCA7 ARE THE CAUSE OF SPINOCEREBELLAR ATAXIA
CC TYPE 7; ALSO KNOWN AS AUTOSOMAL DOMINANT CEREBELLAR ATAXIA TYPE II
CC (ADCA-II). ADCA-II IS CHARACTERIZED BY PIGMENTARY MACULAR
CC DYSTROPHY AND RETINAL DEGENERATION LEADING TO BLINDNESS.
CC -!- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ000517; CAA04154.1; -;
DR EMBL: AF032105; AAC39765.1; -;
DR EMBL: AF032102; AAC19162.1; -;
DR EMBL: AF032103; AAC19163.1; -;
DR MIM: 164500; -;
DR InterPro: IPR000822; Znf_C2H2.
DR SMART: SM00355; Znf_C2H2; 1.
KW Polymorphism; Triplet repeat expansion; Disease mutation;
KW Nuclear protein.
FT DOMAIN 16 20 POLY-ALA.
FT DOMAIN 23 28 POLY-ALA.
FT DOMAIN 30 39 POLY-GLN.
FT DOMAIN 40 45 POLY-PRO.
FT DOMAIN 51 55 POLY-PRO.
FT DOMAIN 171 174 POLY-SER.
FT DOMAIN 213 219 POLY-SER.
FT DOMAIN 647 654 POLY-SER.
FT DOMAIN 717 730 POLY-SER.
FT DOMAIN 840 845 POLY-SER.
FT VARIANT 284 264 K -> R (IN DBSNP:1053338).
FT /FTID=VAR_011823.
FT P -> S (IN DBSNP:1053340).
FT /FTID=VAR_011824.
FT CONFLICT 105 105 P -> H (IN REF. 2; AAC19162).
FT CONFLICT 129 129 C -> S (IN REF. 2; AAC19162).
FT CONFLICT 862 862 V -> M (IN REF. 2; AAC19162).
FT CONFLICT 898 892 PKARP -> VGNG (IN REF. 2; AAC39765 AND
FT AF032103).
SQ SEQUENCE 892 AA; 95451 MW; 9AEA787D77103C5F CRC64;

Query Match 5.3%; Score 96; DB 1; Length 892;
Best Local Similarity 20.1%; Pred. No. 10;
Matches 73; Conservative 45; Mismatches 124; Indels 122; Gaps 16;
QY 17 PPTDSSVTETI-----ILCTMLFLGSLGAWGTTISITGIFSLKTLRSQHGGQVGL 67
Db 567 PSTTSPITRIPHTNSVPTSCQGVSYLAA-----ATVSTFVLLSSTCIS----- 612
QY 68 KYSRPAQPLPAQPPAL-AQPYQSPQPPQTRWVAPRNNAFGSCGAGSDSNSPGNV 126
Db 613 ----PNSKSVPAHGTTLNAQPAASGAMDP-----VCSMQSRQVSSSSSSPST- 655
QY 127 QPNAPSVEHPVL---EKLKAHSYNPKFEFN-----LKSGRVFIKSYSEDD 173
Db 656 -PSGLSSVPSPMRKPQKLKSKSLRPKSSNGSTNCQNASSTSGSGSKKRNKNSPLL 714
QY 174 IHSIKYISLWCTEIGNKRLDFAFR--CMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYG 231
Db 715 VUSSSSSSSSSSSSSSHME---SFRKNCVAHSGPPYP---STVTSSHSIGL----- 758
QY 232 TSAGVNSODKWKGFVDVOWIFVKDVPNNOLR--HRLNNDKPKVYNSRDTQEVPLEKAK 289
Db 759 -----NCVTNKANAVNRHIDQSGRGPTGS-----PAESIK 789
QY 290 OVLKIISYKHTTSI-----FDDFAH----YEKRQRMRMCAKRGVETNN 331
Db 790 RSMVWNSDSTLSLGPFTIHQSNELPVSNGHSFSSHTPLDLKLGKRCSSSSSINNNS 849
QY 332 EGEF 335
Db 850 SSKP 853

```
RESULT 7
MYSC_ACACA STANDARD; PRT; 1168 AA.
AC P10569;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MIC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RT and non-myosin-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86259656; PubMed=3014500;
RA Hammer J.A. III, Jung G., Korn E.D.;
RT "Genetic evidence that Acanthamoeba myosin I is a true myosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
RN [3]
RP PHOSPHORYLATION SITE.
RX MEDLINE=90037074; PubMed=2530230;
RA Bzdeska H., Lynch T.J., Martin B., Korn E.D.;
RT "The localization and sequence of the phosphorylation sites of
RT Acanthamoeba myosin I. An improved method for locating the
RT phosphorylated amino acid.";
RL J. Biol. Chem. 264:19340-19348(1989).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY P-ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
CC VERSA. TH.1 CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
CC OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -1- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J02974; AAA27707.1; -.
CC PIR: A33891; MWAXIC.
CC HSP: P08799; ILVK.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC PRODOM: PD000355; myosin_head; 1.
CC SMART: SM00242; MYSC; 1.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS50002; SH3; 1.
CC MYOSIN: ATP-binding; phosphorylation; Multigene family; SH3 domain.
FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
FT DOMAIN 671 922 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).
```

```
FT DOMAIN 976 1035 SH3.
FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
FT NP_BIND 101 108 ATP (POTENTIAL).
FT MOD_RES 311 311 PHOSPHORYLATION.
SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;

Query Match 5.2%; Score 95; DB 1; Length 1168;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 41; Conservative 20; Mismatches 68; Indels 56; Gaps 7;

QY 45 SISTGSIFSILKTRSQHGQVGLKVRRAOPLPAQPALAOPYOYQSPQ----- 94
Db 857 TLGGKTCVVKFIRDPQGDG--KVKGTKVSVAPCLPSSA-PNIOAQOETSGGASFTVA 913
QY 95 -----PPQTRWVAPRNRAAFQSGGAGSDNSPQNVQPN 130
Db 914 EQSYKQDILGAKGGGGGRGSPSGAVSRP-----SPGGGGGSPSPGGRPSG 968
QY 131 APSVESHVPLEKKAHSY---NPKFEWNLKSRVFIIKSYSEDDIHRSIKYSIWCSTE 187
Db 969 PPAASAPGPEQARALYDFAENPDELTFN--EGAVTVINKSNPD-----WHEGE 1017
QY 188 HGKRR 192
Db 1018 LNGQR 1022

RESULT 8
SOX9_PIG STANDARD; PRT; 509 AA.
AC O18896;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription factor SOX-9.
GN SOX9.
OS Sus scrofa (Pig).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Pilon N., Silversides D.W.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE NORMAL SKELETAL
CC DEVELOPMENT. MAY REGULATE THE EXPRESSION OF OTHER GENES INVOLVED
CC IN CHONDROGENESIS BY ACTING AS A TRANSCRIPTION FACTOR FOR THESE
CC GENES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF029696; AAB81431.1; -.
CC HSP: P48436; ISX9.
CC InterPro: IPR000910; HMG_12_box.
CC Pfam: PF00505; HMG_box; 1.
CC SMART: SM00398; HMG; 1.
CC Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 105 173 HMG BOX.
FT DOMAIN 339 378 GLN/PRO-RICH.
FT DOMAIN 342 349 POLY-PRO.
SQ SEQUENCE 509 AA; 56267 MW; 069F77A0510BBFC2 CRC64;

Query Match 5.2%; Score 94.5; DB 1; Length 509;
```


RT autosomal sex reversal.";
RL Am. J. Hum. Genet. 57:1028-1036(1995).
RN [5]
RX VARIANTS CD LEU-108; ARG-143; PRO-152 AND ARG-170.
RA MEDLINE-97156215; PubMed-9002675;
RA Meyer J., Suedbeck P., Held M., Wagner T., Schmitz M.L.,
RA Bricegelli F.D., Eggermont E., Friedrich U., Haas O.A., Kobelt A.,
RA Leroy J.G., van Maldergem L., Michel E., Mitulla B., Pfeiffer R.A.,
RA Schinzel A., Schmidt H., Scherer G.;
RT "Mutational analysis of the SOX9 gene in campomelic dysplasia and
RT autosomal sex reversal: lack of genotype/phenotype correlations.";
RL Hum. Mol. Genet. 6:91-98(1997).
RN [6]
RX VARIANTS CD SER-112.
RX MEDLINE-98112438; PubMed-9452059;
RA Goji K., Nishijima E., Tsugawa C., Nishio H., Pokharel R.K.,
RA Matsuo M.;
RT "Novel missense mutation in the HMG box of SOX9 gene in a Japanese XY
RT male resulted in campomelic dysplasia and severe defect in
RT masculinization.";
RL Hum. Mutat. Suppl. 1:S114-S116(1998).
RN [7]
RX VARIANTS CD L-112; V-119; Y-165 AND R-170, AND 3D-STRUCTURE MODELING.
RX MEDLINE-99377034; PubMed-1046171;
RA McDowall S., Argentaro A., Ranganathan S., Weller P., Mertin S.,
RA Mansour S., Tolmie J., Harley V.;
RT "Functional and structural studies of wild type SOX9 and mutations
RT causing campomelic dysplasia.";
RL J. Biol. Chem. 274:24023-24030(1999).
RN [8]
RX VARIANTS CD LEU-154 AND THR-158.
RX Preiss A., Argentaro A., Barroso I., Schafer A.J., Clayton A.H.,
RA Ogata T., Harley V.R.;
RT "Functional analysis of two novel SOX9 mutations causing campomelic
RT dysplasia";
RL unpublished observations (JAN-2000).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE NORMAL SKELETAL
CC DEVELOPMENT. MAY REGULATE THE EXPRESSION OF OTHER GENES INVOLVED
CC IN CHONDROGENESIS BY ACTING AS A TRANSCRIPTION FACTOR FOR THESE
CC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: DEFECTS IN SOX9 ARE THE CAUSE OF CAMPOMELIC DYSPLASIA
CC (CD). CD IS A RARE, OFTEN LETHAL, DOMINANTLY INHERITED, CONGENITAL
CC OSTEOCHONDRODYSPLASIA, ASSOCIATED WITH MALE-TO-FEMALE AUTOSOMAL
CC SEX REVERSAL IN TWO-THIRDS OF THE AFFECTED KARYOTYPIC MALES. A
CC DISEASE OF THE NEWBORN CHARACTERIZED BY CONGENITAL BOWING AND
CC ANGULATION OF LONG BONES, UNUSUALLY SMALL SCAPULAE, DEFORMED
CC PELVIS AND SPINE AND A MISSING PAIR OF RIBS. CRANIOFACIAL DEFECTS
CC SUCH AS CLEFT PALATE, MICROGNATHIA, FLAT FACE AND HYPERTELORISM ARE
CC COMMON. VARIOUS DEFECTS OF THE EAR ARE OFTEN EVIDENT, AFFECTING
CC THE COCHLEA, MALLEUS INCUS, STAPES AND TYMPANUM. MOST PATIENTS DIE
CC SOON AFTER BIRTH DUE TO RESPIRATORY DISTRESS WHICH HAS BEEN
CC ATTRIBUTED TO HYPOPLASIA OF THE TRACHEOBRONCHIAL CARTILAGE AND
CC SMALL THORACIC CAGE.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 246629; CAA86598.1; -;
DR EMBL: S74506; AAB32870.1; -;
DR EMBL: S74504; AAB32870.1; JOINED.
DR EMBL: S74505; AAB32870.1; JOINED.
DR PDB: 1SX9; 15-JAN-00.
DR PDB: 1S9M; 03-MAY-00.
DR TRANSFAC: T01853; -;
DR MIM: 114290; -;
DR InterPro: IPR000910; HMG_12_box.

DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; 3D-structure;
KW Disease mutation.
FT DNA_BIND 105 173 HMG_BOX.
FT DOMAIN 339 378 GLN/PRO-RICH.
FT DOMAIN 342 346 POLY-PRO.
FT VARIANT 108 108 P -> L (IN CD).
FT VARIANT 112 112 /FTID=VAR_003735.
FT F -> L (IN CD; LOSS OF DNA BINDING).
FT VARIANT 112 112 /FTID=VAR_003736.
FT F -> S (IN CD).
FT VARIANT 119 119 /FTID=VAR_003737.
FT A -> V (IN CD; ALMOST NO LOSS OF DNA
FT BINDING).
FT VARIANT 143 143 /FTID=VAR_003738.
FT W -> R (IN CD).
FT VARIANT 152 152 /FTID=VAR_003739.
FT R -> P (IN CD).
FT VARIANT 154 154 /FTID=VAR_003740.
FT F -> L (IN CD; 19 FOLD REDUCTION IN DNA
FT BINDING).
FT VARIANT 158 158 /FTID=VAR_008529.
FT A -> T (IN CD; 6 FOLD REDUCTION IN DNA
FT BINDING).
FT VARIANT 165 165 /FTID=VAR_008530.
FT H -> Y (IN CD; LOSS OF DNA BINDING).
FT VARIANT 170 170 /FTID=VAR_008531.
FT P -> R (IN CD).
FT VARIANT 354 356 /FTID=VAR_003741.
FT MISSING (IN CD).
FT /FTID=VAR_003742.
SQ SEQUENCE 509 AA; 56137 MW; 9289CFBB8D6631A2 CRC64;

Query Match 5.1%; Score 93; DB 1; Length 509;
Best Local Similarity 25.2%; Pred. No. 8.5;
Matches 54; Conservative 27; Mismatches 67; Indels 66; Gaps 12;

QY 17 PPTDSSVTFIIILCTMLFLGSLGAWGT--TSISTGSTFSLKTLRSQHGGVGLKVSPPRA 74
Db 307 PATHGVQVTVT-----GSYGISSTATPASAQHW-----MSKQQA 341

QY 75 QPLPAQPPALAQYQSPQPPQTRWVAPRNNAAFQSGGA---GSDSNPGNVQPNFA 131
Db 342 PPPPPQPPQAPP---APQAPPQPPQ--AAPQPPAAPPQPPQAHTLTTLTSSPEPGSQ---- 393

QY 132 PSVESHPVLEKLAH-----SYNPKFEWNLSGRVFIKSYSEDDIHKSIKYSIWCST 186
Db 394 ---RTHIKTEQLSPSHVSEQQHSPQOIAYS-----PFNLPHYSFS--YPPITRSQYDYT 443

QY 187 EHGKNR-----LDSAFRCMS-SKGPVY 207
Db 444 DHONSSSYSHAAGQGTGLYSTETYNPAQRPMT 477

RESULT 12
SYN3_RAT
ID SYN3_RAT STANDARD; PRT; 579 AA.
AC O70441;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Synapsin III.
GN SYN3
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-98256247; PubMed-9593663;

FT DNA_BIND 58 86 MEF2-TYPE (POTENTIAL).
FT DOMAIN 254 257 POLY-PRO.
FT DOMAIN 288 293 POLY-GLU.
FT DOMAIN 419 423 POLY-HIS.
FT DOMAIN 448 455 POLY-SER.
SQ SEQUENCE 498 AA; 53724 MW; 590678D1BD1B3723 CRC64;

Query Match 5.1%; Score 92.5; DB 1; Length 498;
Best Local Similarity 25.7%; Pred. No. 9.1;
Matches 43; Conservative 19; Mismatches 50; Indels 55; Gaps 9;
Qy 15 TSPPTDSSVTETIILCT-----MLFLGSLGAW-----GTTSTIST-----GSIFS 53
Db 332 TAYNTDYSLTSAADLSALQGFSPGMLSLGQASAWQEHHLGGTTLSSLVAGGQLSQGSNLS 391
Qy 54 LKTLRSOHGGGVGLKVSRRPAQLPAQPPALAQ-----PQYQSPQOPQTRWVAPRNR-- 106
Db 392 INTNON-----INIK-SEPISPPRDRWTPSGFQHHHHHPQOQPPQPPQ---PROEMG 442
Qy 107 -----NAAFQGGGAGGSDNS-----PGNVQPNAPSVE 135
Db 443 RSPVDSLSSSSSYDGSREDPRGDFHSPVILGRPANTEDRESFSVK 489

Search completed: July 24, 2002, 19:39:09
Job time: 233 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:34:56 ; Search time 29.22 Seconds
(without alignments)
2012.946 Million cell updates/sec

Title: US-09-877-633-1
Perfect score: 1818
Sequence: 1 MRLGLLOGLTFLRHTSPPTD.....CARNGRVETNNEGPVSYMF 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1307.5	71.9	437	4 Q96S57	Q96S57 homo sapien
2	1307.5	71.9	466	4 Q96AN4	Q96AN4 homo sapien
3	1307.5	71.9	548	4 Q9BTJ7	Q9BTJ7 homo sapien
4	1307.5	71.9	559	4 Q9BYJ9	Q9BYJ9 homo sapien
5	997.5	54.9	570	4 Q9Y5A9	Q9Y5A9 homo sapien
6	978.5	53.8	579	4 Q9BUJ5	Q9BUJ5 homo sapien
7	978.5	53.8	579	11 Q9LYT7	Q9LYT7 mus musculus
8	785.5	43.2	543	4 Q96MZ5	Q96MZ5 homo sapien
9	673.5	37.0	700	5 Q9VBZ5	Q9VBZ5 drosophila
10	600	33.0	197	4 Q9NX79	Q9NX79 homo sapien
11	546	30.0	503	10 Q9L2D8	Q9L2D8 arabadopsis
12	543	29.9	552	10 Q9LVM8	Q9LVM8 arabadopsis
13	521	28.7	493	10 Q9FNR2	Q9FNR2 arabadopsis
14	521	28.7	495	10 Q932P1	Q932P1 arabadopsis
15	521	28.7	528	10 Q9FPE7	Q9FPE7 arabadopsis
16	521	28.7	530	10 O64526	O64526 arabadopsis

17	511.5	28.1	667	10 Q9LJES	Q9LJES arabadopsis
18	507	27.9	549	10 Q94A30	Q94A30 arabadopsis
19	507	27.9	580	10 Q9ZVU7	Q9ZVU7 arabadopsis
20	504	27.7	542	10 Q9C7F6	Q9C7F6 arabadopsis
21	502	27.6	1455	10 Q9LUT8	Q9LUT8 arabadopsis
22	485.5	26.7	664	10 Q9LNG4	Q9LNG4 arabadopsis
23	481.5	26.5	420	10 Q9SQR7	Q9SQR7 arabadopsis
24	465	25.6	708	10 Q94BR4	Q94BR4 oryza sativ
25	464	25.5	499	10 Q9ASM7	Q9ASM7 oryza sativ
26	441	24.3	425	10 Q04503	Q04503 arabadopsis
27	302.5	16.6	306	3 Q06390	Q06390 saccharomyc
28	233.5	12.8	359	10 Q9S262	Q9S262 arabadopsis
29	201.5	11.1	280	5 Q77335	Q77335 plasmodium
30	198.5	10.9	671	4 Q9H6S0	Q9H6S0 homo sapien
31	196.5	10.8	721	5 Q9Y172	Q9Y172 drosophila
32	196.5	10.8	722	5 Q9VZ01	Q9VZ01 drosophila
33	194	10.7	738	11 Q9QY02	Q9QY02 rattus norv
34	188.5	10.4	658	4 Q96M07	Q96M07 homo sapien
35	186	10.2	712	11 Q54729	Q54729 rattus norv
36	113	6.2	437	4 Q96QL2	Q96QL2 homo sapien
37	106.5	5.9	502	10 Q9ZT21	Q9ZT21 triticum ae
38	105	5.8	697	5 Q61121	Q61121 crithidia f
39	105	5.8	931	10 Q9FUY2	Q9FUY2 arabadopsis
40	103.5	5.7	1766	5 Q9V591	Q9V591 drosophila
41	102	5.6	457	5 Q95ZN5	Q95ZN5 caenorhabdi
42	102	5.6	471	2 Q9EZF8	Q9EZF8 streptococ
43	102	5.6	651	5 Q95ZN6	Q95ZN6 caenorhabdi
44	102	5.6	653	5 Q22553	Q22553 caenorhabdi
45	102	5.6	992	4 Q9C0L5	Q9C0L5 homo sapien

ALIGNMENTS

RESULT 1

ID Q96S57 PRELIMINARY; PRT; 437 AA.
AC Q96S57;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DERMATOMYOSITIS ASSOCIATED WITH CANCER PUTATIVE AUTOANTIGEN-1
DE (FRAGMENT).
GN DACA-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Onouchi H., Muro Y., Tomita Y.;
RT "Dermatomyositis Associated with Cancer Autoantigen.*";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055518; BAB62751.1; -.
FT NON_TER 1
SQ SEQUENCE 437 AA; 47281 MW; CA10E70E09BCE47 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 437;
Best Local Similarity 94.6%; Pred. No. 4.5e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY	74	AQPLPAQPALAOPQSQPOPPQTRWVAPRNRNNAFGSGGAGSDSPGNVQPNAPS	133
DB	180	AQPLPAQPALAOPQSQPOPPQTRWVAPRNRNNAFGSGGAGSDSPGNVQPNAPS	239
QY	134	VESHPVLEKLAHSAHYNPKFEWNLKSGRVFIKSYSEDDIHRISKYSTWSTEGHKNRL	193
DB	240	VESHPVLEKLAHSAHYNPKFEWNLKSGRVFIKSYSEDDIHRISKYSTWSTEGHKNRL	299
QY	194	DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQDKWKGFQVWIFV	253
DB	300	DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQDKWKGFQVWIFV	359

QY 254 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 360 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 419
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 314 RQRRRCWCAARNGRVETNNE 332
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

QY 420 RQEEEE-VVRKERQSRNKQ 437
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

RESULT 2

Q96AN4 PRELIMINARY; PRT; 466 AA.
AC Q96AN4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 50.6 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 466;
Best Local Similarity 94.6%; Pred. No. 4.9e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 133
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 209 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 268
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 134 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 269 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 328
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 194 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 329 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 388
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

QY 254 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 389 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 448
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 314 RQRRRCWCAARNGRVETNNE 332
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

QY 449 RQEEEE-VVRKERQSRNKQ 466
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

RESULT 3

Q9BT17 PRELIMINARY; PRT; 548 AA.
AC Q9BT17;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 59.7 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA, CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 548;
Best Local Similarity 94.6%; Pred. No. 6.1e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 133
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 291 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 350
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 134 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 351 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 410
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 194 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 411 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 470
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 254 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 471 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 530
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 314 RQRRRCWCAARNGRVETNNE 332
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

QY 531 RQEEEE-VVRKERQSRNKQ 548
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

RESULT 4

Q9BYJ9 PRELIMINARY; PRT; 559 AA.
AC Q9BYJ9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DJ963E22.1 (NOVEL PROTEIN SIMILAR TO NY-REN-2 ANTIGEN)
DE {FRAGMENT}.
GN DJ963E22.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd D.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL096828; CAC09391.3; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 559 AA; 60873 MW; DB8AA15636130E18 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 559;
Best Local Similarity 94.6%; Pred. No. 6.3e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 133
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 302 AQLPAQPPALAQPOYQSQPPQTRWVAPRNRNAAFQGGAGSDNSPGNVQPNAPS 361
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 134 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 362 VESHPVLEKLAHAHYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 421
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 194 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 422 DSAFCRCSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 481
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

QY 254 KQVNNQLRHRLNNDKPVNTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 548 AA; 59695 MW; B156923EAAAC01DA9 CRC64;

|||||
Db 482 KDVPNNQLRHLRLENNKPVTSRDTQVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 541
QY 314 RQRRRWCARNGRVETNNE 332
Db 542 ROEEEE-VVKKERSRNNQ 559
RESULT 5
Q9Y5A9 PRELIMINARY; PRT; 570 AA.
ID Q9Y5A9
AC Q9Y5A9
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE NGRG8.
GN NGRG8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA Old L.J.;
RT "Antigens recognized by autologous antibody in patients with renal-
cell carcinoma";
RL Int. J. Cancer 0:0-0(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Roberts T.P., Wright A., Wahab N.A., Weston B.S., Mason R.M.;
RT "Gene which is selectively expressed in hyperglycaemia";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF15095; ARD42861.1; -;
DR EMBL; AF192968; AAF08813.1; -;
SQ SEQUENCE 570 AA; 61320 MW; 91FCFA7E508869E4 CRC64;

Query Match 54.9%; Score 997.5; DB 4; Length 570;
Best Local Similarity 75.7%; Pred. No. 4.5e-81;
Matches 190; Conservative 19; Mismatches 35; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAPQYQSPQPPQTRWVAPRNRNAAFGSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQAPQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKFEFENLKSGRVFIKSYSEDDIHRSIKYSIWGSTEHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYSIWGSTEHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPDYCTSGAGVWSQDKWKGFDVQW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGFDVQW 499
QY 251 IFVKDVPNNQLRHLRLENNKPVTSRDTQVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
Db 500 IFVKDVPNSQLRHLRLENNKPVTSRDTQVPLEKAKQVLKIIASYKHTTSIFDDFESH 559
RESULT 6
Q9BUJ5 PRELIMINARY; PRT; 579 AA.
ID Q9BUJ5
AC Q9BUJ5
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE HIGH-GLUCOSE-REGULATED PROTEIN 8.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA, CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002559; AAH02559.1; -;
SQ SEQUENCE 579 AA; 62334 MW; BF3959B5561A464E CRC64;

Query Match 53.8%; Score 978.5; DB 4; Length 579;
Best Local Similarity 75.9%; Pred. No. 2.3e-79;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAPQYQSPQPPQTRWVAPRNRNAAFGSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQAPQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKFEFENLKSGRVFIKSYSEDDIHRSIKYSIWGSTEHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYSIWGSTEHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPDYCTSGAGVWSQDKWKGFDVQW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGFDVQW 499
QY 251 IFVKDVPNNQLRHLRLENNKPVTSRDTQVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
Db 500 IFVKDVPNSQLRHLRLENNKPVTSRDTQVPLEKAKQVLKIIASYKHTTSIFDDFESH 559
QY 311 YEKRQ 315
Db 560 YEKRQ 564

RESULT 7
Q91YT7 PRELIMINARY; PRT; 579 AA.
ID Q91YT7
AC Q91YT7
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 62.3 KDA PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014797; AAH14797.1; -;
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 62280 MW; 9D79DF630593C7AB CRC64;

Query Match 53.8%; Score 978.5; DB 11; Length 579;
Best Local Similarity 75.9%; Pred. No. 2.3e-79;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAPQYQSPQPPQTRWVAPRNRNAAFGSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQAPQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKFEFENLKSGRVFIKSYSEDDIHRSIKYSIWGSTEHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYSIWGSTEHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPDYCTSGAGVWSQDKWKGFDVQW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGFDVQW 499

```
QY 251 IFVKDVPNNQLRHRLNNDNKPVTNRDQEVPLEKAKQVLUKIISYYKHTTSIFDDFAH 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 500 IFVKDVPNSQLRHRLNNDNKPVTNRDQEVPLEKAKQVLUKIISYYKHTTSIFDDF 559
QY 311 YEKRO 315
|||||
Db 560 YEKRO 564

RESULT 8
Q96MZ5 PRELIMINARY: PRT: 543 AA.
AC Q96MZ5;
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE CDNA FLJ131557 FIS, CLONE NT2R12004304, MODERATELY SIMILAR TO HOMO
DE SAPIENS NT-REN-2 ANTIGEN MRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.,
RT *NEDO human cDNA sequencing project.*
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK056219; BAB71122.1; -- 6D3BCE5B83F9BAF5 CRC64;
SQ SEQUENCE 543 AA; 59032 MW; 6D3BCE5B83F9BAF5 CRC64;

Query Match 43.2%; Score 785.5; DB 4; Length 543;
Best Local Similarity 59.9%; Pred. No. 4.5e-62;
Matches 154; Conservative 27; Mismatches 51; Indels 25; Gaps 4;

QY 37 SLGANGTTSISTGSIKTLRSQGGQVGLKYSRRAQP-----LPAQPPALAAQPOY 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 NIGTDW----EKGSYVAKPPTQVLPPTIIQQPOPLIQQPPLVQSLPQPPQPPQ 330
QY 90 QSPQOPP-----QTRWVAPRNRAAFGSGGAGSDSNSPGVQPNAP-SVE 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 OQGPQQAQPHQVQPOQOOLQNRWVAPRNRAAGAGFNQNGAGSENGLGVYVVSASPSSVE 390
QY 136 SHPVLKKAHNSYNPKFEFNKLSGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDS 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 391 VHPVLKKAHNSYNPKFDNKLNGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDA 450
QY 196 APRCMSSGPGVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGWSQDKWKGKPDVQWIFVKD 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 451 AYRSLNGKPLVLLFSVNGSGHFCGVAEMKSYVDYNAVAGWSQDKWKGKFEVVKWIFVKD 510
QY 256 VPNNQLRHRLNNDNKK 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 VPNNQLRHRLNNDNKK 527

RESULT 9
Q9VBZ5 PRELIMINARY: PRT: 700 AA.
AC Q9VBZ5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CG6422 PROTEIN.
GN CG6422.
```

```
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ball J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003750; AAF56381.1; --.
DR FlyBase: FBgn0033261; CG6422.
SQ SEQUENCE 700 AA; 78916 MW; FD1D6C53F769C363 CRC64;

Query Match 37.0%; Score 673.5; DB 5; Length 700;
Best Local Similarity 48.7%; Pred. No. 7.3e-52;
Matches 150; Conservative 35; Mismatches 84; Indels 39; Gaps 9;

QY 70 SRPRAQPL-----PAQPPALAAQFQ-----YQSPQPPQPTRW-----VAPRNRAAFGOS 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 260 SRKYSPTGPVRNLGPPQVPHHAAPRATNAPGPPNARRHDGPHPSRNSERSGNY 319
QY 114 GGAGS-----DSNSPGVQPNAPS---VESHPVLEKKAHNSYNPKFEFNK 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 320 SFRGEFESAKFEYRDENQSRPEATSATELPVDSQLVLDLKDKNYNPKVLD--LK 377
QY 160 ---SGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDSAPRCMSK-KGPVYLLFSVNGS 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 KAGSARFFVIRKSYSEDDIHRSIKYSIMCSTDHGNKRLDDAFKRIEEGGNIMLFFSVNGS 437
QY 216 GHFCGVAEMKSPVDYGTSGAGWSQDKWKGKPDVQWIFVKDVPNNQLRHRLNNDNKPVT 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 438 GHFCGMAQWMTPDVNSTSSYWSQDKWKGKPKVKIYKVDVPGNLTLRHRLNNDNKS 497
QY 276 NSRDQEVPLEKAKQVLUKIISYYKHTTSIFDDFAIYKQRRRRWCAR---NGRVETNNE 332
```


Matches 104; Conservative 43; Mismatches 58; Indels 22; Gaps 4;

```
QY 95 PPOTRWAPNRNNAAFQSGGAGSDSPGNQVPNSAPSVESHVPLEKLAHNSYKPEF 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 302 PRASSRVKSNSSKPCSTIGDSASDSTAG---PN--PSLYNHP-----EF 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 ENWLKSGRVFIKSYEDDIHRSIKYIWCSTHGKRLDSAPR---CMSSKQPVYLLFS 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 343 VTDYKNAKEFFIVKSESDNVHRSIKYVNVWASTPHGNKKLDATAYRDAEKMGGKPIELFS 402
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 VNGSGHFCGVAEMKSPVDYGTAGVWSQDKWKGEKDVQVIFVKDVPNNQLRHRLNNDN 271
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 403 VNASGQFCGVSEMVGPVDFEKDAGYQQDRWSGQFPVKWHIVKIDIPNNRPFCHILLONND 462
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 KPVTSRDTQEVPLEKAKOVKLIISYKHTTSIFDFAHYEKRRQR 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 463 KPVTHSRDSQEVKLRGIEMLRIFREYEARTSILDFGYGYDELEGOK 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 13

```
Q9FNR2 PRELIMINARY; PRT; 493 AA.
AC Q9FNR2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE SIMILARITY TO UNKNOWN PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB066696; BAB10365.1; -.
SQ SEQUENCE 493 AA; 54984 MW; E5539254252DEAF4 CRC64;
```

Query Match 28.7%; Score 521; DB 10; Length 493;
Best Local Similarity 44.9%; Pred. No. 2.2e-38;
Matches 114; Conservative 32; Mismatches 80; Indels 28; Gaps 6;

```
QY 89 YQSPQPPQTRWVA-----PRNRNAAFQSGGAGSDSNSPG-----NVQPNs---APS 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 YDSWKYMP--NWYAVNNTYKPRNGYHCYCKENIEGLNENNRGPRAKGFNSQDGSKYMAVS 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 VESHPVLEKLAH-----SYNPKPEFNNLKSGRVFIKSYSEDDIHRSIKYSIWCS 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 224 LKEQRTVETEKLSQVSLDPKDYNKIDPETYTEAKFYIKSYSEDDIHRSIKYSIWSS 283
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 186 TEHGNKRLDSAPFCMSKSG---PVYLLFSVNGSGHFCGVAEMKSPVDYGTAGVWSQDKW 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 284 TPNGNKKLDASTNEAKQKSGDCCPVLLFSVNTSGQFVGLAEMVGPVDFNKTVYEQDQKW 343
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 KGKFDVQWIFVKDVPNNQLRHRLNNDKPNVTSRDTQEVPLEKAKOVKLIISYKHTT 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 344 ICGFPVKWHFVKDIPNSSLRHITLNNENKPNVTSRDTQEVPLEKAKOVKLIISYKHTT 403
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 SIFDDFAHYEKRRQR 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 404 CILDDFEFENRQK 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 14

Q93ZP1

```
Q93ZP1 PRELIMINARY; PRT; 495 AA.
Q93ZP1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE AT5G61020/MAF19_20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056421; AAL08277.1; -.
SQ SEQUENCE 495 AA; 55211 MW; 55B3EF0C6B6F5CC8 CRC64;
```

Query Match 28.7%; Score 521; DB 10; Length 495;
Best Local Similarity 44.9%; Pred. No. 2.3e-38;
Matches 114; Conservative 32; Mismatches 80; Indels 28; Gaps 6;

```
QY 89 YQSPQPPQTRWVA-----PRNRNAAFQSGGAGSDSNSPG-----NVQPNs---APS 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168 YDSWKYMP--NWYAVNNTYKPRNGYHCYCKENIEGLNENNRGPRAKGFNSQDGSKYMAVS 225
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 VESHPVLEKLAH-----SYNPKPEFNNLKSGRVFIKSYSEDDIHRSIKYSIWCS 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 LKEQRTVETEKLSQVSLDPKDYNKIDPETYTEAKFYIKSYSEDDIHRSIKYSIWSS 285
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 186 TEHGNKRLDSAPFCMSKSG---PVYLLFSVNGSGHFCGVAEMKSPVDYGTAGVWSQDKW 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 TPNGNKKLDASTNEAKQKSGDCCPVLLFSVNTSGQFVGLAEMVGPVDFNKTVYEQDQKW 345
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 KGKFDVQWIFVKDVPNNQLRHRLNNDKPNVTSRDTQEVPLEKAKOVKLIISYKHTT 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 346 ICGFPVKWHFVKDIPNSSLRHITLNNENKPNVTSRDTQEVPLEKAKOVKLIISYKHTT 405
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 SIFDDFAHYEKRRQR 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 406 CILDDFEFENRQK 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 15

```
Q9FPE7 PRELIMINARY; PRT; 528 AA.
AC Q9FPE7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 59.5 KDA PROTEIN.
GN YUP8H12R.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
```



```
SQ SEQUENCE 528 AA; 59549 MW; DF4365FFFE014DF0 CRC64;
```

437 ENKPVTNSRDTQEVRI.POGNEVT.NTEKNYAAKTSTI.DDEDEVENBEK 483

Search completed: July 24, 2002, 19:38:47
Job time: 231 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:32:51 ; Search time 17.76 Seconds
(without alignments)
1839.548 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MRLRGLLQGLTFLRFTSPPTD.....CARNGRVETNNEGEVPSYMF 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138_seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	28.7	530	T01030	hypothetical prote
2	507	27.9	580	C96597	Rubisco subunit bi
3	504	27.7	542	A86405	unknown protein [i
4	441	24.3	425	C86232	hypothetical prote
5	302.5	16.6	306	S61169	hypothetical prote
6	233.5	12.8	359	T06604	hypothetical prote
7	201.5	11.1	280	T18443	hypothetical prote
8	107.5	5.9	1533	A46221	abdominal segment
9	102	5.6	653	T34356	hypothetical prote
10	97.5	5.4	292	T18584	F15C11.2 - Caenorh
11	97.5	5.4	502	T18562	hypothetical prote
12	97.5	5.4	508	S54264	glycoprotein gc -
13	97	5.3	980	S45444	BEM1 protein-bindi
14	96	5.3	367	OZZOZY	circumsporozoite p
15	96	5.3	892	T09193	ataxin 7 - human
16	95	5.2	1168	1MWAYIC	myosin heavy chain
17	94.5	5.2	602	S60052	calcium-dependent
18	94	5.2	520	G86414	probable protein k
19	94	5.2	930	T08588	hypothetical prote
20	93.5	5.1	1074	T24877	hypothetical prote
21	93.5	5.1	1076	T24887	hypothetical prote
22	93	5.1	507	T25831	myocyte-specific e
23	93	5.1	509	A55204	transcription fact
24	93	5.1	1329	A64828	cell division prot
25	93	5.1	2215	T16871	hypothetical prote
26	92.5	5.1	667	T09482	ring finger protei
27	92	5.1	283	T08735	hypothetical prote
28	92	5.1	491	T08477	probable PttA-like
29	92	5.1	491	T12501	hypothetical prote

30 91.5 5.0 645 2 A71416
31 91.5 5.0 785 2 S54016
32 91 5.0 1342 2 G90750
33 91 5.0 1342 2 E85614
34 91 5.0 1353 2 T00347
35 90.5 5.0 741 2 D81798
36 90.5 5.0 877 1 A25962
37 90.5 5.0 1047 2 A59246
38 90.5 5.0 1181 2 C86349
39 90.5 5.0 2515 2 A41519
40 90 5.0 1053 2 T07965
41 89.5 4.9 502 2 T08776
42 89.5 4.9 856 2 T13159
43 89.5 4.9 1460 2 D81675
44 89 4.9 457 2 T04226
45 89 4.9 849 2 S61962

ALIGNMENTS

RESULT 1

T01030

hypothetical protein YUP8H12R.13 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999

C:Accession: T01030

R:rheologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K

Oefner, P.; Davis, R.W.

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A:Reference number: Z14227

A:Accession: T01030

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-530 <PHE>

A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152559; GSPDB:GN00059; ATSP:YU

C:Genetics:

A:Gene: ATSP:YUP8H12R.13

A:Map position: 1

A:Introns: 377/3; 77/1; 100/3; 219/3; 380/3; 451/3; 499/3

Query Match 28.7%; Score 521; DB 2; Length 530;

Best Local Similarity 43.6%; Pred. No. 1.3e-32;

Matches 125; Conservative 28; Mismatches 94; Indels 40; Gaps 7;

QY 61 HGGQVGLKVRPRAPQPLPAQPPALAAQP--QYQSPQQP-----PQTRWVAPRR-- 106

Db 208 HGMGDRPKTPRKASQNSYAPPLNQEKGRIAYPMDPVKKSGALNRDETEKAKARTKEN 267

QY 107 -----NAAFGQ-----SGAGSDDS--NSPGVOPNSAPSVESHVPLEKLAHNSYKPFQEW 156

Db 268 GTSMNDLANGODHITNGECESCSLDAEGNSRNGVSVIRR-----DOYNLPSFQT 318

QY 157 NLKSGRVILKSYSEDDIHSIKYSIWCSTHGKRLDLSAFR-----CMSSGPPVLL 209

Db 319 KYEAIFFVTKSYSEDDIHSIKYNWSSTLNGKKLDSAYOESQKKAADSKGKCPFLP 378

QY 210 FSVNCSGHFCGVAEMKSPVDYTSAGVWSQDKWKGFQVQWIFVKDVPNNLRHRLENN 269

Db 379 FSVNASQFCGVAEMIGRDYDEKSMFEWQDQDKWTGYFPVKWHIHKDVPNQLRHLENN 438

QY 270 DNKPVVNSRDTQEVPLEKAKQVLKLISSYKHTTSIFDDFAHYEKQR 316

Db 439 ENKPVVNSRDTQEVRLPQGVLEVLNIFKNYAARTSILDDDFDYENREK 485

RESULT 2

C96597

Rubisco subunit binding-protein beta subunit [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C96597
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96597
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-580 <STO>
A:Cross-references: GB:AE005173; NID:g4204265; PIDN:AAD10646.1; GSPDB:GN00141
C:Genetics:
A:Gene: TSA14.10
A:Map position: 1

Query Match 27.9%; Score 507; DB 2; Length 580;
Best Local Similarity 35.0%; Pred. No. 1.7e-31;
Matches 122; Conservative 42; Mismatches 113; Indels 72; Gaps 8;

QY 39 GAWGTTSTSGTIFSLKTLRSQGGVGLKVRPRAQ-PLPAQPPALAAQPYQSPQPPQ 97
DB 212 GPTVPVMDHGSNFS-DVORSVSGSVASSYSKANNVPATRNQSSSHYTSYQPAS 270
QY 98 TRWVAPR-----NRNAAPGQ-----SGAGSDSNSPG----- 124
DB 271 MTGAAQGYDVRSPKSYGQGVSTVRSGMGVSGSGYSGTNRGWLNTDNKYRSRGRGN 330
QY 125 -----NV-----QNSAPSVESHVPLEKLA----- 150
DB 331 SYFGNENIDGNEINLRGPRAKTEEVSEEVKKQTFDESNTETVTCVLPDREECN 390
QY 151 PREFEWNKLSGRVFIKSYSEDHHSIKYSYSCSTEGHKNKLSAFRCMSKSG---PVY 207
DB 391 RDEPVEYKDAFFIKSYSEDHHSIKYSYSCSTEGHKNKLSAFRCMSKSG---PVY 450
QY 208 LFFSVNGSHFGCGVAEMKSPDYGTSGAGVSDQKWKGFVDVQIFVKVDPNNOLRHIRE 267
DB 451 LFFSVNASQGTGLAEMKGPVDENKIEYQDQKWTGSPFLKWHILKDVPSNLLKHITLE 510
QY 268 NDNKRPVNSRDTQVPLEKAKQVLISSYKHTTSIFDDFAHYEKROR 316
DB 511 YNENKPVNSRDTQVPLEKAKQVLISSYKHTTSIFDDFAHYEKROR 559

RESULT 3
A86405
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: A86405
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: A86405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <STO>
A:Cross-references: GB:AE005172; NID:g11024874; PIDN:AAG26958.1; GSPDB:GN00141

C:Genetics:
A:Map position: 1

Query Match 27.7%; Score 504; DB 2; Length 542;
Best Local Similarity 43.5%; Pred. No. 2.7e-31;
Matches 108; Conservative 34; Mismatches 68; Indels 38; Gaps 5;

QY 106 RNAAFQSGGA--GSDSNSPG-----NVQPNAPSVESEH-----PVLEKLAHAHSYNPK 152
DB 243 RQAYAAQCRSPFSASSSSPTWENDYNLPLDEARSESYNDPFSHCAPAMLDMLTESNRGPR 302
QY 153 EPEWNLKS-----GRVFIITKSYSEDDHHSIKYSYSCSTEGHKNKLSAFRCMSKSG---PVY 190
DB 303 ASRLNSKSKMISYDVRDRCQOELLSSOFDAKFFVILKSYSEDNHNKHSIKHCVNASTKNGN 362
QY 191 KRLDSAFRCMSKSG---PVYLLFSVNGSHFGCGVAEMKSPDYGTSGAGVSDQKWKGF 247
DB 363 KKLDAAYREAKKKDVACPVFLLFSVNASQFCVAAEMVGPVDFNTSVYEQWQDRWSGHFP 422
QY 248 VOWIFVKVDPNNOLRHIRENNDKPVNTSRTQVPLEKAKQVLISSYKHTTSIFDD 307
DB 423 VOWLIVKDVPSLFRHIIIESNDKPVNTSRTQVPLEKAKQVLISSYKHTTSIFDD 482
QY 308 FAHYEKRQ 315
DB 483 ENFYEBERQ 490

RESULT 4
C86232
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: C86232
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86232
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <STO>
A:Cross-references: GB:AE005172; NID:g2160172; PIDN:AAB60735.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 24.3%; Score 441; DB 2; Length 425;
Best Local Similarity 41.2%; Pred. No. 1.5e-26;
Matches 89; Conservative 35; Mismatches 64; Indels 28; Gaps 3;

QY 132 PSVESHVPLEKLA-----HSPNPKFEFNKLSGRVFIKSYSEDHHSIKYSYSCST 186
DB 136 PPSSESPKONNSFALALRRMYNLPDQTDYEDAKFFVILKSYSEDHHSIKYSYSCST 195
QY 187 EHGKRLDSAFRCMS-----KGPVYLLFS-----VNGSHFGCGVAE 223
DB 196 IGNKKLDAADAEADAEPTKLTLEDCKRRIFLFFSGVLLVCLMSYPSIQVNASQFVGLAE 255
QY 224 MKSPVDYGTSGAGVSDQKWKGFVDVQIFVKVDPNNOLRHIRENNDKPVNTSRTQV 283
DB 256 MVGYVDENKDLDFQWQVQKSGFFPVEHVVKVDIPNWLRLHILDDNEDKPVHTROTTHI 315
QY 284 PLEKAKQVLISSYKHTTSIFDDFAHYEKRORRRR 319

	Query Match	5.6%	Score 102;	DB 2:	Length 653;
	Best Local Similarity	21.7%;	Pred. No. 5;		
	Matches	74;	Conservative	44;	Mismatches 143; Indels 80; Gaps 14;
Qy	12 RFHSPPTDSTVETIILCTMLFLGSLCAWGTTTSISTGSIESLKTLR-----SOHGGOVG	66	: :	:	:
Dd	305 RLRTGVSDLDLE-----DSDDGGGT-----SKWDTYTVRGPRVSADDDGTVR	348	:	: :	:
Qy	67 LKVSRRAPLPALQAOPYOQSPOQPOTRWVPNRNNAFGSQS-GAGSDS----	120	: :	: :	: :
Dd	349 QTRDRPRAO-VDRRSPS--GSPGGTIVRGSPOVAVAEQLRNSSVGYSGGNASAOY	406	:	: :	:
Qy	121 -----NSPCNV-----QPNAPSVEHPVLELKAAHSYNPKFEFWNLKSG---	161	:	: :	:
Dd	407 ATSSLPSHTASSGCATTITLGSENGSPTSLLARTOSVMFSCQRBSAQSWELERGRNP	466	:	: :	:
Qy	162 ---RVFIKYSVEDDIHRSIKYISWCSTEHNKRKLDSAFCRMSSKGPIYLFFSVNGSCH-	217	: :	: :	:
Dd.	467 MSERVRSQQVSPSKYNQHRT-----SSNGVOGGSGGRR-----EYINGSGSG	508	: :	: :	:

A; Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match 5.4%; Score 97.5; DB 2; Length 502;
Best Local Similarity 26.9%; Pred. No. 7.8; Indels 23; Gaps 7;
Matches 45; Conservative 25; Mismatches 74; Indels 23; Gaps 7;

QY 11 LRFHTSPPT-----DSSVTETIILCTMLFLGSLGAMGTTSTSGSIFSLKTLRSQH 61
DB 8 IKVHVKSPTSNDYVEIAADASVSE-LKDKVLVFPVTANKEQVCIIYTKILKDEETLQH 66
QY 62 ---GGVGLKYSRPAQLPAQPPALAAQOYQSPQPPQTRWVAPRNRNNAFGSGGAG 117
DB 67 KIADGHTVHL-VIRNQAARTPA--PAAATPTASSAPSSNPTTSPQNPNTNPFAMGGMG 123
QY 118 SDSNPGNVOPNSVESHVPVLEKLAHAHSYNPKFEWNLKSGRVF 164
DB 124 SPADILNN--PDAMRSVNDNPITQQLLG----NPEFMTIIQSNPQF 164

RESULT 12
S54264
glycoprotein gC - bovine herpesvirus 1
C; Species: bovine herpesvirus 1
C; Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
R; Hecht, P.; Engels, M.; Loeffe, E.; Ackermann, M.
submitted to the EMBL Data Library, May 1995
A; Description: Comparison of the glycoprotein gC genes of bovine and caprine herpesvirus
A; Reference number: S54264
A; Accession: S54264
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-508 <HPC>
A; Cross-references: EMBL:249223; NID:9804966; PIDN:CAA89198.1; PID:9804967
C; Superfamily: herpesvirus glycoprotein F
C; Keywords: glycoprotein

Query Match 5.4%; Score 97.5; DB 2; Length 508;
Best Local Similarity 24.9%; Pred. No. 7.9; Indels 81; Gaps 11;
Matches 57; Conservative 15; Mismatches 76; Indels 81; Gaps 11;

QY 16 SPPTDSSVTEIILCTMLFLGSLGAMGTTSTSGSIFSLKTLRSQHGQVGLKRRPRAQ 75
DB 34 SPPSPSPPTET-----BSSAGTTGATPTTNS-----PDAT 64
QY 76 PLPAQPPALA-----QPQYQSPQPPQTRWVAPRNRNNAFGSGGAGSDNS----- 122
DB 65 PEDSTPGATTGGTPEPPSVSEHDPVTN-STPPPAPPEDGRPGGAGNASRDGRPSGGGR 123
QY 123 PGNVOPNSAPSVESHVPVLEKLAHAHSYNPKFEWNLKSGRVFIILKSYSE-----DD 173
DB 124 PRPQPSKAP-----PDKRWML-CEREAVAASAEPLVYHCGVADN 164
QY 174 IHRSIKYSIMCS-----TEHGKRLDSAFRCMSKGPVYLLFSV-NGS 215
DB 165 ATGCARLELWQVRGRFSTRGCD---DEAVRNPFPRAPPVLLFVVQNGS 210

RESULT 13
S45444
BEM1 protein-binding protein BOB1 - yeast (Saccharomyces cerevisiae)
N; Alternate names: protein YBL0717; protein YBL085W
C; Species: Saccharomyces cerevisiae
C; Date: 09-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
R; Bender, A.; Bender, L.; Kokojan, V.
submitted to the EMBL Data Library, April 1994
A; Description: Yeast Bob1p (Bem1p-binding protein) binds to the SH3 domain-containing pr
A; Reference number: S45444
A; Accession: S45444
A; Molecule type: DNA

A; Residues: 1-980 <BEN>
A; Cross-references: EMBL:L31406; NID:9829041; PIDN:AAB08439.1; PID:9466436
R; Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
submitted to the EMBL Data Library, May 1994
A; Description: Sequence analysis of a 78.6 kb segment of the left end of Saccharomyces
A; Reference number: S45387
A; Accession: S45421
A; Molecule type: DNA
A; Residues: 1-980 <ORE>
A; Cross-references: EMBL:X79489; NID:9496661; PIDN:CAA56021.1; PID:9496694
R; Domdey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.
submitted to the Protein Sequence Database, August 1994
A; Reference number: S45816
A; Accession: S45826
A; Molecule type: DNA
A; Residues: 1-980 <DOM>
A; Cross-references: EMBL:Z35846; NID:9536137; PIDN:CAA84906.1; PID:9536138; GSPDB:GNO
R; Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
Yeast 11, 1103-1112, 1995
A; Title: Sequence analysis of a 78.6 kb segment of the left end of Saccharomyces cere
A; Reference number: S59184; MUID:96076635
A; Accession: S59218
A; Status: nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-980 <OBW>
A; Cross-references: EMBL:X79489; NID:9496661; PIDN:CAA56021.1; PID:9496694
A; Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
C; Genetics:
A; Gene: SGD:BOI1; BOB1; MIPS:YBL085W; SGD:S0000181
A; Cross-references: MIPS:YBL085W; SGD:S0000181
A; Map position: 2L
C; Superfamily: BEM1 protein-binding protein BOB1; pleckstrin repeat homology; SAM hom
F:20-72/Domain: SH3 homology <SH3>
F:225-291/Domain: SAM homology <SAM>

Query Match 5.3%; Score 97; DB 1; Length 980;
Best Local Similarity 19.9%; Pred. No. 21;
Matches 61; Conservative 45; Mismatches 112; Indels 88; Gaps 14;

QY 50 SIFSLSKTLRSQHGQVGLKYSRPAQLPAQPPALAAQOYQSP-----QQPQOT 98
DB 383 NLFADKQIFESPG-----RAPKPPSPSPVQPP-----QSPSFNNRYTNNAFFPQOT 430
QY 99 RWVAPRNRNNAFGSGGAGSDNSPGNVOPNSAPSVESHVPVLEKLAHAHSYNPKFEWNL 158
DB 431 TY-PPKNKPTVYNG-----LIPNSSTSDNSTGKFKPPAMNGHDSNRKTTL 478
QY 159 KSGRVFIKYSSEDI-----HRSIKYSIWCSTEHGKRLDSAFRCMS 201
DB 479 TSATIPSTINTVNTDESPLAISNISSNATSHHPNRNSVYNNHKKRTESGSSFVD-LFNRI 537
QY 202 SKGPVYLLFSYNGSGHPCGVAEMKSPVDYGTSGAGVWSODKWKGRFDVOWIFVKDVPNNQL 261
DB 538 MLSPVKSFDEE-----ETQPSK--ASRAVFDASRRKSSYG-----HSRDASLSEM 582
QY 262 RHIRLENNNDKPVTN--SRDQOEVPLEKAKOVILK-----JSSVKHTTTSIEDD 307
DB 583 K-----KHRRNSILISFFSSKSSQSNPTSPTKQTFTIDPAKMTSHRSRQSNYSYSHARS--QS 636
QY 308 FAHYEK 313
DB 637 YSHSRK 642

RESULT 14
OZZQMY
circumsporozoite protein precursor - Plasmodium yoelii
N; Alternate names: sporozoite surface antigen
C; Species: Plasmodium yoelii
C; Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C; Accession: A26271
R; Lal, A.A.; de la Cruz, V.F.; Welsh, J.A.; Charoenvit, Y.; Maloy, W.L.; McCutchan, T

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:32:01 ; Search time 13.18 Seconds
(without alignments)
630.099 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MRLRGLLQGLRFTSPPTD.....CARNGRVETNNEGEFVSVMF 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1818	100.0	340	2	US-08-933-750C-26
2	1818	100.0	340	4	US-09-234-613-26
3	99	5.4	559	4	US-09-269-731-2
4	96	5.3	478	3	US-08-155-888-2
5	93	5.1	507	5	PCT-US93-08385-8
6	93	5.1	509	4	US-08-860-635A-21
7	93	5.1	509	4	US-09-281-476-21
8	90	5.0	363	3	US-08-586-165-7
9	88	4.8	345	3	US-08-718-738-2
10	88	4.8	345	4	US-09-221-844-2
11	87.5	4.8	580	3	US-08-906-865-1
12	86.5	4.8	521	2	US-08-682-847-4
13	85.5	4.7	188	1	US-08-842-255-131
14	85.5	4.7	198	1	US-08-397-633A-52
15	85.5	4.7	345	5	PCT-US95-03323A-2
16	85	4.7	355	4	US-08-630-915A-192
17	85	4.7	574	4	US-09-276-400-6
18	85	4.7	574	4	US-09-448-076-6
19	85	4.7	623	4	US-09-347-801-4
20	85	4.7	810	6	5200340-8
21	85	4.7	1494	3	US-08-755-587-186
22	84	4.6	339	1	US-08-248-629A-3
23	84	4.6	339	1	US-08-451-932-3
24	84	4.6	339	1	US-08-452-260-3
25	84	4.6	339	1	US-08-326-785-3
26	84	4.6	339	2	US-08-612-788-3
27	84	4.6	339	2	US-08-605-598B-3

Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 40, Appli
Sequence 40, Appli
Sequence 42, Appli
Sequence 42, Appli
Sequence 1, Appli
Sequence 54, Appli
Sequence 54, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 29, Appli

28 84 4.6 339 2 US-08-429-743-3
29 84 4.6 339 2 US-08-866-735-3
30 84 4.6 339 3 US-09-066-028-3
31 84 4.6 339 5 PCT-US95-05107-3
32 84 4.6 352 2 US-08-612-788-40
33 84 4.6 352 3 US-09-066-028-40
34 84 4.6 378 2 US-08-612-788-42
35 84 4.6 378 3 US-09-066-028-42
36 84 4.6 378 4 US-09-206-059-1
37 84 4.6 790 1 US-08-469-486-54
38 84 4.6 790 2 US-08-469-658-54
39 84 4.6 791 1 US-08-643-219-1
40 84 4.6 791 2 US-09-131-995-1
41 84 4.6 791 2 US-08-832-087B-1
42 84 4.6 791 3 US-08-851-350-1
43 84 4.6 791 4 US-09-132-154-1
44 84 4.6 810 1 US-07-854-603-2
45 84 4.6 810 1 US-08-147-000B-29

ALIGNMENTS

RESULT 1
US-08-933-750C-26
; Sequence 26, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNO00003

; CLONE: 1573677
US-08-933-750C-26

Query Match 100.0%; Score 1818; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.3e-173;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLGLLQCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
DB 1 MRLGLLQCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
QY 61 HGGQVGLKYSRRAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
DB 61 HGGQVGLKYSRRAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
QY 121 NSPGNVQPNAPSVEHPVLEKLAHNSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
DB 121 NSPGNVQPNAPSVEHPVLEKLAHNSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
QY 181 SIWCSTEHNKRLDSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGWSQD 240
DB 181 SIWCSTEHNKRLDSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGWSQD 240
QY 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLISSYKH 300
DB 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLISSYKH 300
QY 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340
DB 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340

RESULT 2

US-09-234-613-26
; Sequence 26, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNOBNOT03
; CLONE: 1573677
US-09-234-613-26

Query Match 100.0%; Score 1818; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.3e-173;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLGLLQCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
DB 1 MRLGLLQCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
QY 61 HGGQVGLKYSRRAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
DB 61 HGGQVGLKYSRRAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
QY 121 NSPGNVQPNAPSVEHPVLEKLAHNSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
DB 121 NSPGNVQPNAPSVEHPVLEKLAHNSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
QY 181 SIWCSTEHNKRLDSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGWSQD 240
DB 181 SIWCSTEHNKRLDSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGWSQD 240
QY 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLISSYKH 300
DB 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLISSYKH 300
QY 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340
DB 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340

RESULT 3

US-09-269-731-2
; Sequence 2, Application US/09269731
; Patent No. 6333185
; GENERAL INFORMATION:
; APPLICANT: BARBEYRON, Tristan
; APPLICANT: POTIN, Philippe
; APPLICANT: RICHARD, Christophe
; APPLICANT: HENRISSAT, Bernard
; APPLICANT: YVIN, Jean-Claude
; APPLICANT: KLOAREG, Bernard
; TITLE OF INVENTION: Glycolyse hydrolase genes and their
; TITLE OF INVENTION: use for producing enzymes for the biodegradation of
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DENNISON, MESEROLE, SCHEINER & SCHULTZ
; STREET: 612 Crystal Square 4, 1745 Jefferson Davis
; STREET: Highway
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/269,731

Query Match 5.1%; Score 93; DB 5; Length 507;
Best Local Similarity 25.6%; Pred. No. 0.73;
Matches 44; Conservative 21; Mismatches 49; Indels 58; Gaps 9;

```
OY 15 TSPPTDSSVTETIILCT-----MLFLSLGAW-----GTTISISN-----CSIPS 53
Db 334 TAYNTDYSUTSADLSALQGFNSGMLSLGQVSAWQOHHILGQAALSSLVAGGQLSQGSNLS 393
OY 54 LKTLRSHQGQVGLK---VSRPRAQPLPA-----OPPALAQPQYQSPQPPQPPOT 98
Db 394 INTNQN-----ISIKSEPTSPDRMTFPGFOOQOQOQOQOQPPPPPPQPPQPPQ--PQP 446
OY 99 RWVAPNRNNAFQSGGA---GSDSNS-----PGNVQPNAPSVE 135
Db 447 ROEMGRSPVDSLSSSSSYDSDREDPRGDFHSPVILGRPPNPTEDRESPVK 498
```

RESULT 6

US-08-860-635A-21
; Sequence 21, Application US/08860635A
; Patent No. 6143878

GENERAL INFORMATION:

APPLICANT: Koopman, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,635A
FILING DATE: 29-MAY-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-860-635A-21

Query Match 5.1%; Score 93; DB 4; Length 509;
Best Local Similarity 25.2%; Pred. No. 0.73;
Matches 54; Conservative 27; Mismatches 67; Indels 66; Gaps 12;

```
OY 17 PPTDSSVTETIILCTMLFLSLGAWGT--TSISTGSIFSLKTLRSQHGGQVGLKVSRRPA 74
Db 307 PATHGQVTVT-----GSYGISSTAATPAGAHVW-----MSKQQA 341
OY 75 QPLPAQPPALAAQPOYQSPQPPOTRWVAPNRNNAFQSGGA---GSDSNSPGNVQPN 131
Db 342 PPVPPQPPQPPQPP--APQAPPQPO-AAPQPPQAAAPQPPQAAHTLTITLSSEPGQSG 393
OY 132 PSVESHPVLEKLAH-----SYNPKFEWNLKSGRVFIKSYSEDDIHRSIKYSIWCST 186
Db 394 ---RTHIKTEQLSPSHYSEQQHSPQOIAYS-----PFNLPHYSPS--YPPITRSQDYDT 443
OY 187 EHGKNR-----LDSAFRCMS-SKGPVY 207
Db 444 DHQNSSYVSHAAGQGTGLYSTETYMNPAQRPMY 477
```

RESULT 7

US-09-281-476-21
; Sequence 21, Application US/09281476
; Patent No. 6316597

GENERAL INFORMATION:

APPLICANT: Koopman, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/281,476
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/860,635
FILING DATE:
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-281-476-21

Query Match 5.1%; Score 93; DB 4; Length 509;
Best Local Similarity 25.2%; Pred. No. 0.73;
Matches 54; Conservative 27; Mismatches 67; Indels 66; Gaps 12;

OY 17 PPTDSSVTETIILCTMLFLSLGAWGT--TSISTGSIFSLKTLRSQHGGQVGLKVSRRPA 74

Db 307 PATHQVITYT-----GSGISSTAATPASAGHV-----MSKQOA 341
 Qy 75 QPLPAOPPALAOPVQSPQOPQOTRWAPRNRNNAAGSGGCA---GSDNSPGNVQPNMSA 131
 Db 342 PPPPPQPPQAPP---APQAPPOPO-AAPPOQAPPOQAPQAHLLTLLSSEPGOSQ-----393
 Qy 132 PSVESHVPYLEKKAH-----SYNPKFEFENLKSGRVFIKSYSEDDTHRSIKYSIMCST 186
 Db 394 ---RTHIKTEOLSFESHSEOOHSPQOLAYS-----FENLPHYSPPS---YPPITRSQDYDT 443
 Qy 187 EHGKNR-----LDSAFRCMS-SKGPVY 207
 Db 444 DHQNSSSYISHAAGCGTGLYSTFTYMNPAQRPMY 477

RESULT 8

US-08-586-165-7

; Sequence 7, Application US/08586165
 ; Patent No. 6054298
 ; GENERAL INFORMATION:
 ; APPLICANT: Laufer, Edward M.
 ; APPLICANT: Orozco, Olivia E.
 ; APPLICANT: Tabin, Clifford J.
 ; TITLE OF INVENTION: Fringe Proteins and Pattern Formation
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: US
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/586.165
 ; FILING DATE: 16-JAN-1996
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Granahan, Patricia
 ; REGISTRATION NUMBER: 32,227
 ; REFERENCE/DOCKET NUMBER: H095-05
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 861-6240
 ; TELEFAX: (617) 861-9540
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 363 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-586-165-7

Query Match 5.08; Score 90; DB 3; Length 363;

Best Local Similarity 20.68; Pred. No. 0.87; Mismatches 142; Indels 104; Gaps 19;

Qy 22 SVTETIILCTMLFL-----GSLG-AWGTTSTIGTSIFSLKTLRSQHGQVGL-----K 68
 Db 12 SLVGSMTCLLVLMVEPPGRGLARGEAGGAORALQSLGAARACQAGPGLRTADYGR 71
 Qy 69 VSRPRAQPLAOPPALAOPVQSPQOPQ-----TRWV 101
 Db 72 LSRARRE-LPAAPP-----SPRPPAEDITPRDVFIAVKTTKFKHAKLELLLDTWI 122
 Qy 102 APRNRNNAAGSGGAGSD-SNSPGNVQNSAPSVEHPVLEKLAHSAHYNPKFEWNLKS 160
 Db 123 S-RNRDNTFTFDGRDELKQKARNVINTNCSAAHSROALSKMAV-----EYDKFTIES 175

Qy 161 GRVFIKSYSEDDTHRSIK--YSIWCSTEH-----GNKRLDSAPRCM-----SKSGPVY 207
 Db 176 GRKWF--CHVDDDNVNVVRLTKLLSSYPHTQDIYIGRPSLDRLPQOATERISENKMHPVH 233
 Qy 208 LLFSVNSGSHFC---GVAEMKSPVDYG-----TSAGVNSQDKWKGFQVQVIFVKDVPNN 259
 Db 234 FWEATGGAG-FCISRGLALKMSPWASGGHFMSTAERILPDDCTTIGYIESVLGVKLIRS 292
 Qy 260 QLRHRLNNDNKPVTNSPDQEVPLEKAKQV---LKIISYKHTTSTFDDFAHVEKQR 316
 Db 293 NLFHSHLENLHOVPKT-----EIHQVTLVSYGMFNKRNSIHKMGAFSVEEDPSR 342
 Qy 317 RRR-----WCARN 324
 Db 343 FRSVHCLLYPDTWPCPSN 360

RESULT 9

US-08-718-738-2

; Sequence 2, Application US/08718738
 ; Patent No. 6013469
 ; GENERAL INFORMATION:
 ; APPLICANT: KUNSCH, CHARLES A
 ; APPLICANT: CHOPRA, ARVIND
 ; APPLICANT: ROSEN, CRAIG A
 ; TITLE OF INVENTION: HUMAN B-CELL TRANSLOCATION GENES-2 AND 3
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/718.738
 ; FILING DATE: 18-SEP-1996
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEFFE, ERIC K
 ; REGISTRATION NUMBER: 36,688
 ; REFERENCE/DOCKET NUMBER: 1488.0730003/EKS/KMT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 345 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-718-738-2

Query Match 4.88; Score 88; DB 3; Length 345;

Best Local Similarity 27.28; Pred. No. 1.3; Mismatches 16; Indels 8; Gaps 4;

Qy 46 ISTGSIFSLKTLRSQHGQVGLKV---SRPRAQPLPAQPPALAAQYQSPQPPQTRWA 102
 Db 211 LNVNDLLKQAKAISSMHSLYGLGLGSGQQQOQOQAPQPPPPPPQOQOQO--KTSALS 268
 Qy 103 PRNRNNAAGSGGAGSDNS--PGNVQNSAPSVEHPVLEKLAHSAHYNPKEF 154
 Db 269 PNAKEFITPNMOGOGSSTNGMFGDSPLNLSPLQYSN-AFDVFAAYGGLNEKSF 321

```
RESULT 10
US-09-221-844-2
: Sequence 2, Application US/09221844
: Patent No. 6258777
: GENERAL INFORMATION:
: APPLICANT: KUNSCH, CHARLES A
: APPLICANT: CHOPRA, ARVIND
: APPLICANT: ROSEN, CRAIG A
: TITLE OF INVENTION: HUMAN B-CELL TRANSLOCATION GENES-2 AND 3
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/221,844
: FILING DATE: 29-DEC-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/718,738
: FILING DATE: 18-SEP-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,382
: FILING DATE: 05-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/460,104
: FILING DATE: 02-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/03323
: FILING DATE: 17-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: MILLONIG, ROBERT C
: REGISTRATION NUMBER: 34,395
: REFERENCE/DOCKET NUMBER: 1488.0730004/EKS/RCM/SCW
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 345 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-221-844-2

Query Match 4.8%; Score 88; DB 4; Length 345;
Best Local Similarity 27.2%; Pred. No. 1.3;
Matches 31; Conservative 16; Mismatches 59; Indels 8; Gaps 4;

QY 46 ISFGSIFSLKTRSGHGGVGLKV---SRPRAQLPAAQPPALAAQYQSPQPPOTRWVA 102
Db 211 LNVNDLKKAKISSHSHSYGLGLGSGQQPQQQQQPAQPPPPPPPPPPPPPPPPPPPPPP 268
QY 103 PRNRNAAFGSGGAGSDSNS--PGNVQPSAPSVESHPVLEKLAHNSYNPKF 154
Db 269 PNAKEFIFFNMGGSTNGMFGDPLNLSPLQYSN-AFDVFAAYGGLNEKSF 321

RESULT 11
US-08-906-865-1
: Sequence 1, Application US/08906865
: Patent No. 6040168
: GENERAL INFORMATION:
: APPLICANT: Greengard, Paul
```

```
: APPLICANT: Porton, Barbara
: APPLICANT: Kao, Hung-Teh
: TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David A. Jackson, Esq.
: STREET: 411 Hackensack Ave, Continental Plaza, 4th
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/906,865
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-202
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-487-5800
: TELEFAX: 201-343-1684
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 580 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: DESCRIPTION: /desc = "Synapsin III"
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
US-08-906-865-1

Query Match 4.8%; Score 87.5; DB 3; Length 580;
Best Local Similarity 30.0%; Pred. No. 3.1;
Matches 30; Conservative 18; Mismatches 49; Indels 3; Gaps 3;

QY 54 LKTLRSQHGGVGLKYSRPAQLPAAQPPALAAQ-POYQSPQPPQTRWVAPRNRNNAFGQ 112
Db 414 IKSAKSPGAQLGLPQLXQPPRPPOGGPRQAQSPQPSQOR-LSPOGQQPLSPQ 472
QY 113 SGGAGSDSNSPGNVQPSAPSVESHPVLEKLAHNSYNPK 152
Db 473 S-GSPQQRSPGSPQLSRASSGSSPNQASKPGATLASQPR 511

RESULT 12
US-08-682-847-4
: Sequence 4, Application US/08682847
: Patent No. 5858989
: GENERAL INFORMATION:
: APPLICANT: BABIUK, LORNE
: APPLICANT: VAN DEN HURK, SYLVIA
: APPLICANT: ZAMB, TIM
: APPLICANT: FITZPATRICK, DAVID
: TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 PAGE MILL ROAD
: CITY: PALO ALTO
: STATE: CA
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:28:11 ; Search time 33.33 Seconds
(without alignments)
1133.066 Million cell updates/sec

Title: US-09-877-633-1
Perfect score: 1818
Sequence: 1 MRLRGLLQGLRFTSPPTD.....CARGRVETNNEGPVSYMF 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	991.5	54.5	570	AA39510	Human polypeptide
2	978.5	53.8	564	AA41296	Human polypeptide
3	978.5	53.8	579	AA36626	Human FLEXIT-48 pr
4	673.5	37.0	700	AB62329	Drosophila melanog
5	521	28.7	429	AA42527	Arabidopsis thalia
6	521	28.7	511	AA42526	Arabidopsis thalia
7	521	28.7	530	AA42525	Arabidopsis thalia
8	516	28.4	315	AA08665	Arabidopsis thalia
9	515	28.3	257	AA46038	Arabidopsis thalia
10	515	28.3	260	AA08667	Arabidopsis thalia
11	515	28.3	286	AA46037	Arabidopsis thalia

12	515	28.3	289	21	AA08666	Arabidopsis thalia
13	507	27.9	292	21	AA08666	Arabidopsis thalia
14	507	27.9	293	21	AA08666	Arabidopsis thalia
15	507	27.9	294	21	AA08666	Arabidopsis thalia
16	507	27.9	295	21	AA08666	Arabidopsis thalia
17	506.5	27.9	274	21	AA08666	Arabidopsis thalia
18	506.5	27.9	275	21	AA08666	Arabidopsis thalia
19	504	27.7	461	21	AA08666	Arabidopsis thalia
20	504	27.7	465	21	AA08666	Arabidopsis thalia
21	504	27.7	503	21	AA08666	Arabidopsis thalia
22	502	27.6	529	21	AA08666	Arabidopsis thalia
23	502	27.6	1421	21	AA08666	Arabidopsis thalia
24	502	27.6	1499	21	AA08666	Arabidopsis thalia
25	502	27.6	1655	21	AA08666	Arabidopsis thalia
26	496	27.3	493	21	AA08666	Arabidopsis thalia
27	495.5	27.3	471	21	AA08666	Arabidopsis thalia
28	438.5	24.1	326	21	AA08666	Arabidopsis thalia
29	438.5	24.1	355	21	AA08666	Arabidopsis thalia
30	438.5	24.1	388	21	AA08666	Arabidopsis thalia
31	339	18.6	277	21	AA08666	Arabidopsis thalia
32	277	15.2	256	21	AA08666	Arabidopsis thalia
33	217	11.9	45	22	AA08666	Human NY-REN-2 Ag
34	197.5	10.9	407	20	AA08666	Renal cancer assoc
35	196.5	10.8	722	22	AA08666	Drosophila melanog
36	187.5	10.3	712	18	AA08666	Rat YTS21 gene pro
37	179	9.8	274	21	AA08666	Arabidopsis thalia
38	131	7.2	312	22	AA08666	Human NY-REN-2 Ag
39	105	5.8	480	21	AA08666	Human protein tran
40	103.5	5.7	1766	22	AA08666	Drosophila melanog
41	103	5.7	839	21	AA08666	Human prostate can
42	101	5.6	578	22	AA08666	Human colon cancer
43	99	5.4	559	19	AA08666	Alteromonas fortis
44	98.5	5.4	786	22	AA08666	Novel human diagno
45	98.5	5.4	786	22	AA08666	Novel human diagno

ALIGNMENTS

RESULT 1		
ID	AA39510	standard; Protein; 570 AA.
XX	AA39510;	
AC	AA39510;	
XX	AA39510;	
DT	22-OCT-2001	(first entry)
XX	Human polypeptide	SEQ ID NO 2655.
DE	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	peripheral nervous system; neuropathic; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;	
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukaemia.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	WO200153312-A1.	
PN	26-JUL-2001.	
XX	26-JUL-2001.	
PD	26-DEC-2000; 2000WO-US34263.	
XX	26-DEC-2000; 2000WO-US34263.	
PF	21-JAN-2000; 2000US-0488725.	
XX	25-APR-2000; 2000US-052317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
PR	29-NOV-2000; 2000US-0727344.	
XX		

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB: AAI58666.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4: SEQ ID NO 2655; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 570 AA;

Query Match 54.5%; Score 991.5; DB 22; Length 570;
Best Local Similarity 75.3%; Pred. No. 2.2e-89;
Matches 189; Conservative 19; Mismatches 36; Indels 7; Gaps 4;
QY 73 RAQPLPAQPPALAAQPOQYQSPQPPQTRWVAPRNRNAAFGCGAGSDNSPGNVQ--PNS 130
DB 325 qtqplppppppaqqlsvqqaaqp-trwvaprnrsgfghn---gvdngvgsgqsgsgs 380
QY 131 APSVESHVPLEKLAHNSYNPKFEWNLSGSRVFIKSYSEDDIHRSTKYSTWCSTEHGN 190
DB 381 tps-ephvpvleklrsinnynpkdfwnlkhgrvfiiksyseddihrsikynlwcstehgn 439
QY 191 KRLSAFCRMSKSGPVYLLFVSGSGHFCGVAEMKSPVDYGTFSAGVWSQDKWKFDFVQW 250
DB 440 krldaayrsmngkpgpyllfsvngsgghfcgvaemksavdyntcagvwsqdkwkgfdrvrw 499
QY 251 IFVKDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISYKHTTSIFDDFAH 310
DB 500 ifvkdvpnsqrlrhlrlnndnkpvtnsrtdtqevplekakqvlkllasykhttsifddfh 559
QY 311 YEKRRRRRNC 321
DB 560 yekrprgrkrc 570

RESULT 2
AAM41296
ID AAM41296 standard; Protein; 564 AA.
XX
XX AAM41296;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 6227.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS Homo sapiens.
XX WO200153312-A1.
PN 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 19-JUL-2000; 2000US-0598042.
PR 03-AUG-2000; 2000US-0620312.
PR 14-SEP-2000; 2000US-0653450.
PR 19-OCT-2000; 2000US-0662191.
PR 29-NOV-2000; 2000US-0693036.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB: AAI60452.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6227; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 564 AA;

Query Match 53.8%; Score 978.5; DB 22; Length 564;
Best Local Similarity 75.9%; Pred. No. 4.2e-88;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;
QY 73 RAQPLPAQPPALAAQPOQYQSPQPPQTRWVAPRNRNAAFGCGAGSDNSPGNVQ--PNS 130
DB 325 qtqplppppppaqqlsvqqaaqp-trwvaprnrsgfghn---gvdngvgsgqsgsgs 380
QY 131 APSVESHVPLEKLAHNSYNPKFEWNLSGSRVFIKSYSEDDIHRSTKYSTWCSTEHGN 190
DB 381 tps-ephvpvleklrsinnynpkdfwnlkhgrvfiiksyseddihrsikynlwcstehgn 439
QY 191 KRLSAFCRMSKSGPVYLLFVSGSGHFCGVAEMKSPVDYGTFSAGVWSQDKWKFDFVQW 250
DB 440 krldaayrsmngkpgpyllfsvngsgghfcgvaemksavdyntcagvwsqdkwkgfdrvrw 499
QY 251 IFVKDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISYKHTTSIFDDFAH 310
DB 500 ifvkdvpnsqrlrhlrlnndnkpvtnsrtdtqevplekakqvlkllasykhttsifddfh 559

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 28.7%; Score 521; DB 21; Length 429;
Best Local Similarity 43.6%; Pred. No. 7.6e-43;
Matches 125; Conservative 28; Mismatches 94; Indels 40; Gaps 7;

QY 61 HGGQVGLKVSRRAPRAQPLPAOPPALAQP--OYQSPQQP-----PQTRWVAPRR-- 106
DB 107 hmgdrpkrpkasqsyappllnqekgrlaypmdpkkksalnrdetkarkten 166
QY 107 -----NAAFGQ----SGGAGSDS--NSFGNVQPNAPSVEHPVLEKLKAAHSYNPKFEW 156
DB 167 gtsmdliangqdhntgecescsldaegnersngvgvirr-----dqynlpsft 217
QY 157 NLKSGRVPIIKSYSEDDIHRISIKYISWCSTPHGNKRLDSAPR-----CMSSKGPVYLL 209
DB 218 kyecalfvfkysceddhksikynvwsstlngnkikidsaygesqkkaadkskpcvflf 277
QY 210 FSVNGSGHFCGVAEMKSPVDYGTSGAGVMSODKWKGFQDVOWIFVKDVFNOLRHLRLNN 269
DB 278 fsvnasgqfcgvaemigrvdyeksmefwqdkwcyfpvkkwhlikdvpnpqirhlenn 337
QY 270 DNKPTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYERQR 316
DB 338 enkptnsrtdtqevrlpggnevlnfknyaaktsildddfdyenrek 384

RESULT 6
AAG42526
ID AAG42526 standard; Protein; 511 AA.
XX
AC AAG42526;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53047.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0123845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143544.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161922.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match			28.7%; Score 521; DB 21; Length 511;
Best Local Similarity			43.6%; Pred. No. 9.9e-43;
Matches 125; Conservative			28; Mismatches 94; Indels 40; Gaps 7;
QY	61	HGGVGLKVRPRAPQPLPAQPPALAAQP--QYQSPQQP-----PQTRWVAPRNR--	106
Db	189	hgmdrptktrkasqnsyapppplnqkgriaypmdpvkksqalnrdekekakartken	248
QY	107	-----NAAFQG-----SGAGSDS--NSPGNVQPNAPSVESHPVLEKLUKAHSYNPKFEF	156
Db	249	gtamndlangqdhntgcescsldaegnersngsvir-----dqynlpsfqt	299
QY	157	NLKSGRVFIITKSYSEDDIHSIKYISWCSSTEHGNKRLDLSAFR-----CMSSKGPVYLL	209
Db	300	kyeaaifvikiyseddihksikynvwsstlngnkkidsaygesqkkaadskgkcpvlf	359
QY	210	FSVNGSCHFCGVAEMKSPVDYGTSGAGVSDKWKGKFDVQWIFVKDVPNNQLRHRLENN	269
Db	360	fsvnasgfcgvaemigrvdyeksmetwqgdkwtgyfpvkwhiikdvppqlrhille	419
QY	270	DNKPVNTRDTEVPLEKAKOVLKLISSYKHTTSIFDDEFAHYEKKROR	316
Db	420	enkpvtnsrdtqvrlbpqgnevinifknyaaktsilddffeyenrek	466
RESULT 7			
AAG42525			
ID	AAG42525 standard; Protein; 530 AA.		
XX	AAG42525;		
AC	AAG42525;		
DT	18-OCT-2000 (first entry)		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 53046.		
DE	Arabidopsis thaliana		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX	Arabidopsis thaliana.		
OS	Arabidopsis thaliana.		
XX	EP1033405-A2.		
PN	EP1033405-A2.		

XX	06-SEP-2000.	
PD		
XX		
PE	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	25-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 6294.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0133426.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 28.3%; Score 515; DB 21; Length 286;
Best Local Similarity 51.7%; Pred. No. 1,6e-42;
Matches 105; Conservative 22; Mismatches 60; Indels 16; Gaps 2;

```

Db 51 daegnersngvgsvrr-----dqynlpsftkyeeaaiffviksyseddihksaiky 101
QY 181 SIMCSTEHGKRLDSAFR-----CMSSKGPVYLLFSVNGSGHFVCGVAENKSPVDYGTG 233
Db 102 nvsstingnkkldsaygesqkkaadskgkcpvflfsvnasgfcygaemigrvdyeks 161
QY 234 AGVNSQDKWKGKFDVQWIFVKDVPNNOLRHIRLENNDNKPVTSNRDQEVPLEKAKQVLK 293
Db 162 mefqqqdkwtgyfvpkwhiikdvnpqlrhiilennenkpvtnsrtdqevrlpqagnevln 221
QY 294 IISYKHTTSIFDDFAHYERQR 316
Db 222 ifknyaaktsildddfyenrek 244

RESULT 12
AAG08666
ID AAG08666 standard; Protein; 289 AA.
XX
AC AAG08666;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6295.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
```

```
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155639.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 28.3%; Score 515; DB 21; Length 289;
Best Local Similarity 51.7%; Pred. No. 1.6e-42;
Matches 105; Conservative 22; Mismatches 60; Indels 16; Gaps 2;

QY 121 NSFGNVPNSAPSVESHPVLEKLAHAHSYNPKFEPENLKSGRVFIKSYSEDDIHSIXY 180
DB 51 daegnersngvsgvirr-----dqynlpsfqkyceaffvfkysyseddihsiky 101
QY 181 SJWCSTEHCNKRLDASFR-----CMSSKGPVYLLFSVNGCHFCGVAEMKSPVDYGT 233
DB 102 nwsstlngnkldasayqesqkkaadksqcpvlfsvnasggqfcvgaemigrvdyeks 161
QY 234 AGVWSQDKWKGFQVQWIEFKVDPVNNQLRHIRELNNNDKRPVTNSRDTQVPLEKAKQV 293
DB 162 mefwqdkwtgyfvpkwhilkdvppqlrhillelennenkpvtnsrdtqevr!pggnev 221
QY 294 IISYKHTTSIFDDFANYEKRR 316
DB 222 lfknvaaktsilddfdfyenrek 244

RESULT 13
AAG21588
ID AAG21588 standard; Protein; 292 AA.
XX
AC AAG21588;
XX
DT 17-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24198.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
```


PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159395.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.


```
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.9%; Score 507; DB 21; Length 293;
Best Local Similarity 50.0%; Pred. No. 1e-41;
Matches 103; Conservative 27; Mismatches 62; Indels 14; Gaps 3;

Qy 125 NVQPNs---APSVESHVPVLEKLAHAH-----SYNPKFEWNLKSGRVFIKSYSDDD 173
Db 11 nsqdgskvmavslkeqrvtetekisedvslldpkdyndkidfpetyteakfyvikysedd 70

Qy 174 IHRSTIKYSIMCSTHGKRLDSAFRCMSKG---PVYLLFSVNGSGHFCGVAEMKSPVDY 230
Db 71 ihksikysvwsstpngnkkldasyneakqksdgcpcvflfsvntsqgfvglamvgpvdv 130

Qy 231 GTSAGVNSQDKWKCKFDYQWIFVKDPNQLRHIRENNDKPVTNSRDTQVPLEKAKQ 290
Db 131 nktveywqgdkwigcfvkvwhfvkdpnsslrhitlennenkpvtnsrtdgcvkleqgik 190

Qy 291 VLKTISSYKHTTSIPDDFAHYEKQR 316
Db 191 vikifkhashtcllddfeftenrk 216

RESULT 15
AAG21587
ID AAG21587 standard; Protein; 294 AA.
XX
AC AAG21587;
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 24197.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
```

PR 06-MAY-1999; 99US-0134487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0134487.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.9% Score 507; DB 21; Length 294;
Best Local Similarity 50.0%; Pred. No. 1e-41;
Matches 103; Conservative 27; Mismatches 62; Indels 14; Gaps 3;

QY 125 NVQPN-----APSVESHVPVLEKLAH-----SYNPKFEFWNLKSGRVFIKSYSEDD 173
Db 13 nsqdgskvmavslkeqrvtetekisedvsllldpkdynkidfpetyteakfyviksysedd 72

QY 174 IHRSIKYSINCSHGKRLDSAPRCMSSKG---PVYLLFSVNGSGHFCGVAEMKSPVDY 230
Db 73 ihksikyvvwsstpgnkkidasyneakqksdgcpcvflfsvntsggfvglaeenvgpvdf 132

QY 231 GTSAGVMSODKWKGFVDVONTFVVDVNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKO 290
Db 133 nktveyqqdkwlgcfpkywhfvkdiposslrhlennnenkpvtnsrtdgevkiedgik 192

QY 291 VLKIISYKHTTSIFDDFAHYEKRQR 316
Db 193 vikikdhashtcilddfefyenrqk 218
```

Search completed: July 24, 2002, 19:34:54
Job time: 403 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 16:44:49 ; Search time 237.28 Seconds
(without alignments)
14674.235 Million cell updates/sec

Title: US-09-877-633-2
Perfect score: 2028
Sequence: 1 caaaaggacaagataataaa.....agtagacagagtgatgga 2028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736435 seqs, 858457221 residues 3472872

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1773	87.4	3043	22	AAH72969 Human cervical can
2	1534.6	75.7	2685	22	ABA08628 Human NY-REN-2 Ag
3	870	42.9	8382	22	AAK70731 Human immune/hacma
4	341.6	16.8	1718	20	AA217776 Human gene express
5	341.6	16.8	2750	22	AAK88117 Human FLEXHT-48 nu
6	341.6	16.8	3029	22	AA160452 Human polynucleoti
7	340.6	16.8	3845	22	AA158666 Human polynucleoti
8	226	11.1	379	22	AA158666 Novel human polynu
9	225	11.1	226	14	AA061419 Human brain Expres

c	10	218.4	10.8	625	20	AAK90820	Rat progression re
	11	213	10.5	747	22	AAI95660	Human neuroblastom
	12	209.6	10.3	2218	23	ABL11025	Drosophila melanog
c	13	195.4	9.6	6461	23	ABL11024	Drosophila melanog
c	14	173.2	8.5	307	22	AAI08157	Human breast cance
c	15	173.2	8.5	307	22	AAI19067	Human breast cance
c	16	171.4	8.5	293	22	AAI25902	Human breast cance
	17	167.4	8.3	1514	20	AAK39643	Renal cancer assoc
	18	158	7.8	711	20	AAZ15119	Human gene express
	19	144.4	7.1	950	21	AAK44533	Human secreted exp
	20	139.4	6.9	1679	21	AAK32849	Arabidopsis thalia
	21	136.6	6.7	1135	21	AAK39320	Arabidopsis thalia
	22	136.6	6.7	1136	21	AAK46031	Arabidopsis thalia
	23	135.4	6.7	2675	14	AAO41914	Soybean seed beta-
	24	133.8	6.6	779	22	AAI97361	Human neuroblastom
	25	132.4	6.5	861	21	AAK48546	Arabidopsis thalia
	26	132.4	6.5	1209	21	AAK34339	Arabidopsis thalia
	27	132.4	6.5	1849	21	AAK47231	Arabidopsis thalia
	28	122.8	6.1	739	22	AAI95659	Human neuroblastom
	29	122.2	6.0	729	20	AAV87783	EST clone ES306.
	30	120.2	5.9	1853	21	AAK45480	Arabidopsis thalia
	31	119.4	5.9	1854	21	AAK38976	Arabidopsis thalia
	32	119.4	5.9	4968	21	AAK46149	Arabidopsis thalia
	33	114.8	5.7	925	20	AAV88879	EST clone HZ103.
	34	112	5.5	464	21	AAA42435	Human secreted exp
c	35	103	5.1	400	22	AAK25444	Human ovarian PCR-
	36	101.8	5.0	1586	21	AAK36395	Arabidopsis thalia
	37	94.8	4.7	1749	22	ABA08809	Human NY-REN-2 Ag
	38	94.4	4.7	488	22	ABA09487	Human renal cancer
	39	92	4.5	460	21	AAA43413	Rat secreted expre
	40	84.4	4.2	214	21	AAK07592	Human secreted pro
	41	82.4	4.1	246	21	AAK07593	Human secreted pro
	42	66.4	3.3	399	22	AAK67243	Novel human polynu
	43	65.4	3.2	14006	24	ABL32958	Human immune syste
	44	64.8	3.2	3261	22	AAF26571	DNA encoding human
	45	63.8	3.1	6145	24	ABL32972	Human immune syste

ALIGNMENTS

RESULT 1

AAH72969

ID AAH72969 standard; cDNA: 3043 BP.

XX AC AAH72969;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 4243.

XX DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX KW Homo sapiens.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0169681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

XX DR

QY 759 ggacagcgctccgctgcattgacgacgaaggccgcgtctactctcttccagctcaa 818
|||||
Db 1227 ggacagcgctccgctgcattgacgacgaaggccgcgtctactctcttccagctcaa 1286
QY 819 tgggaagtgggcatttttgggggtggccgagatgaagtcccccgtagactacgacccag 878
|||||
Db 1287 tgggaagtgggcatttttgggggtggccgagatgaagtcccccgtagactacgacccag 1346
QY 879 tgcgcgggtctggtctcaggaacaagtgggaagggaagtgtgatgtccagtggaatttttgt 938
|||||
Db 1347 tgcgcgggtctggtctcaggaacaagtgggaagggaagtgtgatgtccagtggaatttttgt 1406
QY 939 taaggatgtaccocaaataccagctccggccacatcagcctggagaaatacagacaacaacc 998
|||||
Db 1407 taaggatgtaccocaaataccagctccggccacatcagcctggagaaatacagacaacaacc 1466
QY 999 ggtcacaaactccggggacacccagaggtgccccttagaaaaagccaagcagctgtaa 1058
|||||
Db 1467 ggtcacaaactccggggacacccagaggtgccccttagaaaaagccaagcagctgtaa 1526
QY 1059 aattatcagttctctacaagcacacacacccctccatcttctgcgaccttctgctcacctacagaaa 1118
|||||
Db 1527 aattatcagttctctacaagcacacacacccctccatcttctgcgaccttctgctcacctacagaaa 1586
QY 1119 gcgcca-gaggaggaggaggtggtgcgaagggaacgcgagagtcgnaaacaacaatgagg 1177
|||||
Db 1587 gcgcccaggaggaggaggtggtgcgaagggaacgcgagagtcgnaaacaacaatgagg 1646
QY 1178 gggaaacagatttctctacagttcttaacgittgactttgaaacacagttttaaacaacagtgtg 1237
|||||
Db 1647 gggaaacagatttctctacagttcttaacgittgactttgaaacacagttttaaacaacagtgtg 1706
QY 1238 ctgtggtcagctccagctgtgtcgtcccgctgcgggggttgagtggtgcattcttgcctttct 1297
|||||
Db 1707 ctgtggtcagctccagctgtgtcgtcccgctgcgggggttgagtggtgcattcttgcctttct 1766
QY 1298 tgtcgttgatttttggcccgatggatctgcattattttgtacttttttctatgtattataa 1357
|||||
Db 1767 tgtcgttgatttttggcccgatggatctgcattattttgtacttttttctatgtattataa 1826
QY 1358 tctgttagaagtcacataataaaggagatttttttttgcagcttctcaatcagactgat 1417
|||||
Db 1827 tctgttagaagtcacataataaaggagta-ttttttttgcagcttctcaatcagactgat 1885
QY 1418 cttaatgtgaattgaagtatcccttaaaaaacaagcactatttttggcagaaaattgtgttc 1477
|||||
Db 1886 cttaatgtgaattgaagtatcccttaaaaaacaagcactatttttggcagaaaattgtgttc 1945
QY 1478 ttaaatctcagtcatttgatattctgtgagacttcatttctcattccctttattgttttt 1537
|||||
Db 1946 ttaaatctcagtcatttgatattctgtgagacttcatttctcattccctttattgttttt 2005
QY 1538 tagcaacataagaacccatgacttcttgcatttgcatttagagattctgataaactcttt 1597
|||||
Db 2006 tagcaacataagaacccatgacttcttgcatttgcatttagagattctgataaactcttt 2065
QY 1598 gaaaaactagtaaaactaaaagggttaagtatttttgcatttgcatttgcatttttatta 1657
|||||
Db 2066 gaaaaactagtaaaactaaaagggttaagtatttttgcatttgcatttgcatttttatta 2125
QY 1658 tctgttatcgggttaagtgcataatttaacatttgcatttttctgcgtagacagataaactt 1717
|||||
Db 2126 tctgttatcgggttaagtgcataatttaacatttgcatttttctgcgtagacagataaactt 2185
QY 1718 ttaatttttcaaatctggcagacactttttttttttttttttttttgaaaactttctccagat 1777
|||||
Db 2186 ttaatttttcaaatctggcagacacttttttttttttttttttttttggaaaactttctccagat 2245
QY 1778 ctg-ttggccactgaacagcca-cccgctccctcactgtcctggtgtccgattgggtgga 1835
|||||
Db 2246 ctgttttggccactgaacagccaccgcctccctcactgtcctggtgtccgattgggtgga 2305
QY 1836 tgggtgtggggcattgattgtgtggaggaaactgggaaggtgcttttaggtctggttcagggtcg 1895

|||||
Db 2306 tgggtgtggggcattgattgtgtggaggaaactgggaaggctgttttaggtctggttcagggtcg 2365
QY 1896 ggcatctcttggttgttgcacatctttttaaaattttacacattttcttaagaatt---cta 1953
|||||
Db 2366 ggcatctcttggttgttgcacatctttttaaaattttacacattttcttaagaatttctaa 2425
QY 1954 atccctctttaaagttttttataccaataatgctgagctttaaagtg-aggatctcgggtagta 2012
|||||
Db 2426 atccctctttaaagttttttataccaataatgctgagctttaaagtgtaaggatctcgggtagta 2485
QY 2013 cagacagtgatgagg 2027
|||||
Db 2486 cagacagtgatgagg 2500
RESULT 3
AAK70731/C
ID AAK70731 standard; DNA; 8382 BP.
XX
AC AAK70731;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:25543.
XX
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
PD
PP 09-AUG-2001.
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.

PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249219.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249224.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232081.	PR	17-NOV-2000;	2000US-0249300.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251859.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	XX		
PR	27-SEP-2000;	2000US-0235834.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235836.	XX		
PR	29-SEP-2000;	2000US-0236327.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	XX		
PR	29-SEP-2000;	2000US-0236367.	DR	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	02-OCT-2000;	2000US-0236802.	PT	metastasis.	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	PS	Disclosure; SEQ ID NO 25543; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
PR	13-OCT-2000;	2000US-0239937.	CC	amino acid sequences given in AAM621170 to AAM91921. (I) have cytosstatic	
PR	20-OCT-2000;	2000US-0240960.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	20-OCT-2000;	2000US-0241221.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-0241785.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241785.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241786.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241787.	CC	that affect the activity of (I) by expressing inactive proteins or to	
PR	20-OCT-2000;	2000US-0241808.	CC	supplement the patients own production of (I). Additionally, (I)	
PR	20-OCT-2000;	2000US-0241809.	CC	the nucleic acids into a host cell and culturing the cell to express the	
PR	20-OCT-2000;	2000US-0241826.	CC	protein. (I) proteins and polynucleotides may be used to prevent,	
PR	01-NOV-2000;	2000US-0244617.	CC	diagnose and treat immune/haematopoietic-related diseases, especially	
PR	08-NOV-2000;	2000US-0246474.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
PR	08-NOV-2000;	2000US-0246475.	CC	AAK67694 represent human immune/haematopoietic antigen genomic	
PR	08-NOV-2000;	2000US-0246476.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAM62169	
PR	08-NOV-2000;	2000US-0246477.	CC	represent sequences used in the exemplification of the present invention.	
PR	08-NOV-2000;	2000US-0246478.	CC		
PR	08-NOV-2000;	2000US-0246523.	XX		
PR	08-NOV-2000;	2000US-0246524.	SQ	Sequence 8382 BP; 2411 A; 2122 C; 1929 G; 1920 T; 0 other;	
PR	08-NOV-2000;	2000US-0246524.			
PR	08-NOV-2000;	2000US-0246525.			
PR	08-NOV-2000;	2000US-0246526.			
PR	08-NOV-2000;	2000US-0246527.			
PR	08-NOV-2000;	2000US-0246528.			
PR	08-NOV-2000;	2000US-0246532.			
PR	08-NOV-2000;	2000US-0246533.			
PR	08-NOV-2000;	2000US-0246609.			
PR	08-NOV-2000;	2000US-0246610.			
PR	08-NOV-2000;	2000US-0246611.			
PR	08-NOV-2000;	2000US-0246611.			
PR	17-NOV-2000;	2000US-0249207.			
PR	17-NOV-2000;	2000US-0249208.			
PR	17-NOV-2000;	2000US-0249209.			
PR	17-NOV-2000;	2000US-0249210.			
PR	17-NOV-2000;	2000US-0249211.			
PR	17-NOV-2000;	2000US-0249211.			

Query Match 42.9%; Score 870; DB 22; Length 8382;

Best Local Similarity 99.9%; Pred. No. 1.5e-216;

Matches 881; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1147	aggaacgcgcagagtcgaaacaaacatgagggcgacacgtttcttaccatgtcttaacgt	1206
Db	1308	AGGAACGGCAGAGTCGAAACAAACAAATGAGGGGCAACCACTTCTTACATGTTCTAAAGT	1249
QY	1207	ttgactttgaaacacagtttaaacacgctgtgcttggtcagctccagtgatgtctcccg	1266
Db	1248	TTGACTTTGAAACACAGTTTAAACACAGTGTGCTTGTCAGTCACAGTGTGTCGCTCGG	1189


```
QY 716 agtactccatctgtgtagcacagacagcgcaacacagcgccctggacagcgccctccgct 775
Db 418 agtataatattgtgcagcacagacatggttaacagagactggtctgtctatcggt 477
QY 776 gcatgagacagcaaggcgccctacactgctcttcacgctcaatggagtgagcatttt 835
Db 478 ccatgaacgggaaggcccgcttacttactcttcagtgctcaacggcagtgagcactct 537
QY 836 gtggggtggccgagatgaagtcgcccgctggactacgacagcgccacagtgccgggtctggctc 895
Db 538 gtggcggtggcagaaatgaactctgctggactacacacacatgacagtgctggctcc 597
QY 896 aggcacagtggaagggaagttgatctccagtgagttttttgtaaggatgacccaata 955
Db 598 aggcacaaatgaagggtcgttttgatgctcagtgagttttttgtaaggagcttcccaata 657
QY 956 accagctccggcacatcaggtcgagataacgacacacacacacacggttcacaaactcccg 1015
Db 658 gccactgacacattcgcctagacacacagcagagataacacacagtgaccacactctg 717
QY 1016 acaccaggaaggtgcccttagaagaaagcgaagcagtgctgaaatattatcagttcctaca 1075
Db 718 acactcaggaagtcctctgaaaggctagcaggtgttgaaattatagccagctaca 777
QY 1076 agcacacacccatcttcagacacttgcctactacgagagcgcc-agagagagagg 1134
Db 778 agcacacacactctattttgatgactctcactatgagaaacgccaagaggaag 837
QY 1135 aggtggtgcgaaggaacgacagagtcgaacacacaa 1172
Db 838 aaagtgttaaaaggagcgtcaaggtcgctgggaaataa 875
```

RESULT 5

AAC88117

ID AAC88117 standard; cDNA; 2750 BP.

XX AC AAC88117:

XX DT 09-MAR-2001 (first entry)

XX XX Human FLEXHT-48 nucleotide sequence SEQ ID NO:103.

XX KW Human; FLEXHT; full-length molecules expressed in human tissue;
XX KW diagnosis; gene expression; genetic linkage; genetic variability;
XX KW antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
XX KW cytosolic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
XX KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
XX KW antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;
XX KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
XX KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;
XX KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
XX KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
XX KW ulcerative colitis; ss.

XX OS Homo sapiens.

XX PN WO200070047-A2.

XX XX 23-NOV-2000.

XX XX 12-MAY-2000; 2000WO-US13299.

XX XX 14-MAY-1999; 99US-0311894.

XX PR 14-MAY-1999; 99US-0311937.

XX PR 14-MAY-1999; 99US-0311940.

XX XX (INCY-) INCYTE GENOMICS INC.

XX PA

XX PI Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;

XX PI Azimzal Y, Lu DAM, Au-Young J, Shih LL;

XX XX WPI; 2001-016234/02.

DR P-PSDB; AAB36626..

XX Human FLEXHT protein and DNA sequences, useful for treating
PT immunological disorders, developmental disorders, and cancers -

XX Claim 5; page 163-164; 168pp; English.

XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules
CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
CC present invention describes an isolated polypeptide (A) comprising an
CC amino acid sequence selected from one of 55 amino acid sequences 42-876
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
CC identical sequence, and a biologically active or immunogenic fragment of
CC the sequence. The FLEXHT proteins can have antianemic, anticonvulsant,
CC antiarteriosclerotic, immunomodulatory, cytosolic, antisthmatic,
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,
CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antiulcer
CC and antirheumatic activities, and can be used in gene therapy. The
CC polynucleotide sequences can be used to express the protein sequences.
CC Pharmaceutical compositions comprising FLEXHT can be used to treat
CC diseases or conditions associated with altered expression of functional
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
CC treat disorders including anaemia, epilepsy, arteriosclerosis,
CC atherosclerosis, developmental disorders, cancers, and immunological
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
XX ulcerative colitis.

SQ Sequence 2750 BP: 776 A; 619 C; 638 G; 717 T; 0 other;

Query Match

16.8%; Score 341.6; DB 22; Length 2750;

Best Local Similarity

67.5%; Pred. No. le-78;

Matches 512; Conservative 0; Mismatches 239; Indels 7; Gaps 2;

QY 416 cacagccccagcttggctcaacgcgagtcagagccctcagcagcagccccagacc 475

Db 1162 ctccacctccaccagcgtgccagctttcagtcacgacacagcagctcagcaacc 1221

QY 476 gctgggtggccacgcaacagaaacggcggtttggcagagcgaggggtggcagcg 535

Db 1222 gctgggtgagcctcggaacccgtggcagtggtgctcagtaataatgggtggtg 1281

QY 536 atagcaactctctggaaacgccagctcctctgccccagcgtcgatcccccccg 595

Db 1282 gagtaggacagctcagcgtggtctggtgactctctcttcttcttcttcttctt 1335

QY 596 tcttgaaacctgaagctgctcagctcagctcagctcagctcagctcagctcagctc 655

Db 1336 tgttggaagcttcggtccattataataataataataataataataataataata 1395

QY 656 aaagcggtggtgttctatcatcaagagctactctgagagacacatccacgctcc 715

Db 1396 aacatggcggtgttctatcatcaagagctactctgagagacacatccacgctcc 1455

QY 716 agtactcctctggttagcacagacgagcaacacgctcgtgacagcgcttccgct 775

Db 1456 agtaataatttggcgagcagacagcagcagcagcagcagcagcagcagcagcag 1515

QY 776 gcatgagcagcaagggtggtggtggtggtggtggtggtggtggtggtggtggt 835

Db 1516 ccatgaacgggaaggccccgtttacttacttacttacttacttacttacttactt 1575

QY 836 gtgggtggccgagatgaagtcgcccgctggactacgagcagcagcagcagcagcag 895

Db 1576 gtggcgtgagaaatgaaatctgctggactacacacacacacacacacacacac 1635

QY 896 aggcacagtggaagggaagtttgatgctcagtgatgttttttgaaggatgaccca 955

Db 1636 aggcacaaatgaagggtgtgtgtgagtggtgtgtgtgtgtgtgtgtgtgtgtgt 1695

QY 956 accagctcggcagacatcaggtgggagataacgacacacacacacacacacacac 1015

Db 1696 gccactgcgacacattccctagagaaacagagagataaacacagtgaccacactcaggg 1755
Qy 1016 acaccagagaggctccttagaaaaagcccaagcaagtgcgaaattatcagttcctaca 1075
Db 1756 acactcaggaagcgctctggaagggctaaagcagggctgaaattatagccagctaca 1815
Qy 1076 agcacaaacctccattcttcgacgactttgctcactacgagaagcgcc-agaggaagagg 1134
Db 1816 agcacaccattccattttttgatgacttctcacactatgagaacgccaagagaagaag 1875
Qy 1135 agtggtgcgcaggaagcagcgagagtcgaaacaaacaa 1172
Db 1876 aaagtgttaaaaggaagcagtcgaagtcgtgggaataaa 1913

RESULT 6
AAI60452/c
ID AAI60452 standard; cDNA; 3029 BP.
XX
AC AAI60452;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4441.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0485725.
PR 05-APR-2000; 2000US-0552317.
PR 19-JUL-2000; 2000US-0598042.
PR 03-AUG-2000; 2000US-0620312.
PR 14-SEP-2000; 2000US-0653450.
PR 19-OCT-2000; 2000US-0662191.
PR 29-NOV-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41296.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4441; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 3029 BP; 766 A; 737 C; 715 G; 809 T; 2 Other;

Query Match 16.8%; Score 341.6; DB 22; Length 3029;
Best Local Similarity 67.5%; Pred. NO. 1.1e-78;
Matches 512; Conservative 0; Mismatches 239; Indels 7; Gaps 2;

Qy 416 cacagccccagcttggctcaaccgcagtcagagccctcagcagcccccagacc 475
Db 1593 CTCACCTCCACCACAGCTGCCAGCTTTCAGTCGACCAACAGGACGTCAGCCAACCC 1534
Qy 476 gctgggtgccccacgcacacagaaacgcggtttggcagagcgaggggctggcagcg 535
Db 1533 GCTGGGTAGCACCTCGGAACCCGTGGCAGTGGGTTCGGTCAATATGGGTGGTAAATG 1474
Qy 536 atagcaactctcctggaaaacgtccagcctaatcttgcctccacgcgtgcgaatcccccccg 595
Db 1473 GAGTAGGACAGTCTCAGGCTGGTTCGTGGATCTACTCTCTTC-----AGAACCCACCCAG 1420
Qy 596 tccttaaaaaaacgaaggtgctcacagctcacacccgaaagagtttgatgaactcga 655
Db 1419 TGTGGAGAGCTTCGCGTCCATTAACTATTAACCCCAAGATTTTGTACTGGAATCTGA 1360
Qy 656 aaagcggggctggttcacatcaatcaagagctactctgagagcagacatccacgcctcatta 715
Db 1359 AACTGCGCGGTTTCATCATTAAGAGCTACTCTCAGGAGGATATTCACCGTTCCATT 1300
Qy 716 agtactccatcttggttagcacagagcaagcgcaagcgccctggacagcgcttcocgt 775
Db 1299 AGTATAATATTTGGTGCAGCACAGCATGGTAACAAGAGACTGCGATGCTTATCGTT 1240
Qy 776 gcatgagcagcaagggcccgctctcctctcctcctcctcctcctcctcctcctcctcct 835
Db 1239 CCATGAACGGGAAGGCCCGCTTACTTACTTTTCACTAGTGTCAACGGCAGTGGACACTCT 1180
Qy 836 gtgggggtggccgagatgaagtcctcccggtgactacagcagcagtcgctgggtctcctc 895
Db 1179 GTGGCGTGGCAGAAATGAATCTGCTGGACTTACAACACATCTGCAGAGTGTGTGGTCCC 1120
Qy 896 aggacaagtggaaaggggaagtttgatgctccagtggttttcttaagggatgtaccacaata 955
Db 1119 AGGACAAATGGAAGGGTCTGTTTGTATGTCAGTGGGATTTTGTGAAGGAGCTTCCCAATA 1060
Qy 956 accagctccggcacatcaggtgagaaataacacacacacacacacacacacacacacac 1015
Db 1059 GCCAACTGGACACATTCGCCTAGAGAACCAACGAGATAAACCCAGTACCAACTCTAGGG 1000
Qy 1016 acaccagaggtgctccttagaaaaagcgaagcgaagtgctgaaataatcagttcctaca 1075
Db 999 ACACCTCAGGAAGTGCCTCTGGAAAAGCTTAAGCAGGTGTGNAATATATAGCCAGCTACA 940
Qy 1076 agcacacacctccatctcttcgacgacttttgcctacacagagagcgcc-agaggaagagg 1134
Db 939 AGCACACCACTTCCCATTTTGTATGACTTCTCACACTATGAGAAACGCCAAGAGGAAGAAG 880
Qy 1135 aggtggtgcgcaggaagcagcgaggtcgaaacaaacaa 1172
Db 879 AAGAGTGTAAAAAGGAACGCTCAAGCTCGTGGGAATAA 842

RESULT 7
AAI58666
ID AAI58666 standard; cDNA; 3845 BP.
XX
AC AAI58666;

[illegible]

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX WPI: 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
XX Claim 9: Page 928; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
XX Sequence 379 BP; 100 A; 80 C; 95 G; 103 T; 1 other;

Query Match 11.1%; Score 226; DB 22; Length 379;
Best Local Similarity 75.7%; Pred. No. 5.7e-49;
Matches 280; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 583 gaatccaccacccgctcgtgaaacacgaagctgtcagctacacacccgaaagagttt 642
Db 9 gaaccaccacccagtggtgagaagcttcggtccattataactataaccccaagatttt 68
QY 643 gagtgaatctaaaaagcgggtgtgttcacatcaacagactactctgaggacgacatc 702
Db 69 gactggaatctgaacacatggccgggttttcacatcaacagactactctgaggacgata 128
QY 703 caccgtccatcaagactccactgtgtgtgtagcacagacgacggaacacgctggagac 762
Db 129 caccgtccatcaagataataatttgggtgcagcacagacgacatggtaacagagactggat 188
QY 763 agcgcttcctgcgtgacgacgacgaaggggcccgctctacactgctcttcagcgtcaatggg 822
Db 189 gctgcttatcgttcctcatgaacgggaaaggcccgcttactactcttcagtgctcaacgac 248
QY 823 agtgggcatcttttggtgggtgcccagagatgaagtcaccccgctggactacggccacagtcgc 882
Db 249 adtgacacatctctgtggcgtggcagaatgaaatctgctgtgagactacacacatgtgca 308
QY 883 ggggtctgtctcagacacaaagtgggaagggaattgtgatgtccagtggaattttgttaag 942
Db 309 ggtgtgtgtgtcccccagacaaaatgggaagggttcgtttcgtgctcaggtggatttttggaag 368
QY 943 gatgtaccaca 952
Db 369 gacgttccca 378

RESULT 9
AAO61419/c
ID AAO61419 standard; DNA; 226 BP.
XX
XX AAO61419;
XX
DT 16-MAR-1994 (first entry)

XX Human brain Expressed Sequence Tag EST01406.
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX Homo sapiens.
XX
XX WO9316178-A.
XX
XX 19-AUG-1993.
XX
XX 12-FEB-1993; 93WO-US01772.
XX
XX 12-FEB-1992; 92US-0837195.
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Adams MD, Moreno RF, Venter CJ;
XX WPI: 1993-272882/34.
XX
XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
XX Example 4: Page 488; 500pp; English.
XX
XX The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST01406 has a "marginal" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
XX Sequence 226 BP; 81 A; 53 C; 46 G; 45 T; 1 other;
XX
Query Match 11.1%; Score 225; DB 14; Length 226;
Best Local Similarity 99.6%; Pred. No. 7.8e-49;
Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1757 taaaaatcttctccacatctgttgcacacgacacccctccactctct 1816
Db 226 TGAATAATCTTCTCCAGATCTGTTGCCACTGACACGCCCCCTCCCTACTCTCT 167
QY 1817 ggtgtccgattgggtggatggtgttggggcatgctgtggaggaactggaggtgctt 1876
Db 166 GGTGTCCGATTGGGCTGGATGCTGTTGGGCGCATGATGTGTGGAGGAANTGGAAGTCTT 107
QY 1877 taggtctggttcaggctcgggcatcttctgttggatcatcttttttaattttacac 1936
Db 106 TAGGTCTGTTTACGGTCGGGCATTTCTTTGTTGTTTGCACATCTTTTAAATTTTACACC 47
QY 1937 ttttctaaagaattcaatgcgctcttaagttttttatataccaataat 1982
Db 46 TTTTCTTAAAGATTCCTAATGCCGCTCTTAAGTTTATATACCAATAAT 1
XX
RESULT 10
AAQ90820
ID AAX90820 standard; cDNA; 625 BP.
XX
XX AAX90820;
XX
XX 13-JAN-2000 (first entry)
XX
XX Rat progression related cDNA, ZNPE-62-1 5'end partial sequence.
XX
XX Progression related gene; PRG; ZNPE-62-1; rat; progression; diagnosis;
KW

QY 804 gctcttcagcgtcaatggagtgggca--tctttgtgggtggtccgagagatgaagtcctccc 861
II IIIII IIIII II II IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 562 ACTTTTTCAGTGTCAACGGCCAGTGGACACTTCTTGTGGCGTGGCAGAAATGAAATCTGCT 503
QY 862 gtgactacggcacacagtcgcgggtctgtctctcaggacaaagtgaagggtttgat 921
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 502 GTGGACTACAACACATGTCAGGTGTGTGTGCCAGG-CAAATGGAAGGTCGTTTGTAT 444
QY 922 gtccagtgattttgttaagatgtaccataaccagctccggcacatcaggtggag 981
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 443 GTACAGTGGATTTTGTGAAGGACGTTCCCAATAGCCAATCGCACACATTCGCCCTAGAG 384
QY 982 aatacagacacaaacccggtcacaactcccgggacacccaggaggtgccccttagaaaaa 1041
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 383 AACACAGAGAATAAACCACTGACCAACTACTAGGACACTCAGGAAGTGCCTTGGAAAG 324
QY 1042 gccagcagtgctgaaattatcagttctctacagcacacacacccctctcttcagcag 1101
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 323 GCTAAGCAGGTGTTGAAAATTATAGCCAGTACAGCACACCACTTCCATTTTGTATGAC 264
QY 1102 ttgtctcactacagaaagcgc--agagagagagaggtgtgtgcgaagaaagcagagt 1160
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 263 TTCTACACTATGAGAACCCCAAGAGGAGGAAGAAGTGTAAAAGGACGTCAAGGT 204
QY 1161 cgaacacaaacaa 1172
II IIIII
Db 203 CGTGGGAATAA 192
RESULT 12
ID ABL11025 standard; cDNA; 2218 BP.
XX ABL11025;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27557.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB66922.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Claim 1; SEQ ID NO 27557; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2218 BP; 689 A; 571 C; 535 G; 423 T; 0 other;
Query Match 10.3%; Score 209.6; DB 23; Length 2218;
Best Local Similarity 63.8%; Pred. No. 2.8e-44;
Matches 352; Conservative 0; Mismatches 194; Indels 6; Gaps 2;
QY 580 gtggaactgcacccctccttgaaaaactgaaggctgtccacagctcacaccccgaaagag 639
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1175 gtgattcccaactcgtgttgacgaattaaaaagacaaacaattacaaccaaagtg 1234
QY 640 ttgagtggtg--aatctgaaaagcggcggtgtgttcatcatcaagagctactctgaggac 696
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1235 ttggacttgaaaaagccggatcgccagattcttctgtcatcaaatcgtactccgaagt 1294
QY 697 gacatcacgcgtccatgaagtaactcctctgtgtgtagcacagacacggcaacaagcgc 756
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1295 gacatccatcgcaagtataaagtatgagatttgggtcctcaaccgcatcgcgcaataagaga 1354
QY 757 ctggacagcgccttcctcgctgcatgagcagcaag--gggcccgtctacctgctcttcagc 813
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1355 ttggacgatgccttcaaggagcgaacgaaggcggcgcaacataatgctgtctctctcg 1414
QY 814 gtcaatgggagtggtggtatgttgggtggtggcgagatgaagtcctcccgctggactacgac 873
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1415 gtgaatggatctggtccattctcggcgtatggcacaatgatgacgcggtcgactacaac 1474
QY 874 accagtcgggggtctgtctcaggacaaagtgaagggaagtttcatgtccagtgatt 933
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1475 tcaacgtcagcgtctgtgtcgcaaggagagtggtggaggtccaggtcgaaatggata 1534
QY 934 ttgttaagatgtaccacaataaacacagctccggcacatcaggtctgagataaacgcacac 993
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1535 tatgtcaagatgtgcttaacgggaacgtcgtcaccatcgcactggagaacaatggaat 1594
QY 994 aaacgggtcacaaactcccgggacacccggaggtgccccttagaaaaagcgaagcgaagt 1053
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1595 aagtcggtgacgaactcaaggggacacccaggaagtgcccaacgacaaagcgcaggtg 1654
QY 1054 ctgaaaattatcagttcctacaaagcacaacctccatctcgaagctttgtctcactac 1113
II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 1655 ctgcagattttgcactcgtacaaaccattcaacgtccattcgtatgacttctccactat 1714
QY 1114 gagaagcgcag 1125
II IIIII
Db 1715 gaaaagaagcag 1726
RESULT 13
ID ABL11024/c
XX ABL11024 standard; cDNA; 6461 BP.
XX
XX ABL11024;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27554.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX

PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB66921.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 27554; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6461 BP; 1715 A; 1377 C; 1462 G; 1907 T; 0 other;
SQ

Query Match 9.6%; Score 195.4; DB 23; Length 6461;
Best Local Similarity 65.8%; Pred. No. 2.6e-40;
Matches 302; Conservative 0; Mismatches 156; Indels 3; Gaps 1;
QY 668 ttttcatcatcaagagctactctgaggacacatccaccgctccatttaagtactccatct 727
DB 2017 TCITTTGTCATCAATCTACTCCGAAGATGACATCCATCCAGTATTAAGTATGAGATT 1958
QY 728 ggtgtagcacagagcacggcaaacagcgcttgagcagcgcttcctcgatgagcagca 787
DB 1957 GGTGCTCAACCGATCAGCGGAATAAGAGATTGGACGATGCGCTTCAAGGAGCGCACGAAG 1898
QY 788 aggggcccgtctac---ctgctcttcagcgctcaaatggagtgggcatttttgggggtg 844
DB 1897 AGGCGGCAACATAATGCTGTCTTCGTGGTGAATGGATCTGGCCATTTCTGCGGCATGG 1838
QY 845 ccgagatgaagtcctccgtgactacgacacagtcgagcggtctgtctcagagacaagt 904
DB 1837 CACAATGATGATGACCGCGGTGACTTACAACTCAAGCTCGAGCGTCTGGTCGAGGACAAAGT 1778
QY 905 ggaaggggaagtgtgatgtccagtggtatttttggtaaggatgtaccacaataaccagctcc 964
DB 1777 GGAGGGGCAAGTTCAAGGTGAATGGATATATGTCAAAGATGTGCTTAACGGAACGCTGC 1718
QY 965 ggcacatcaggctgagaaataacagacaacaaaccgggtcacaaaactccgggagacccagg 1024
DB 1717 GTCACATTCGACTGGGAACAATAAGATAAGTCGGTGACGAATCAAGGGACACCCAGG 1658
QY 1025 aggtgccttagaagaaagcaagcaagtgcgtgaaattatcagttcctacaagcacaaa 1084
DB 1657 AAGTGGCCCAACGACAAGGCAATCAGGCTGCTGCAGATTTTGCACCTCGTACAAACCATTCAA 1598
QY 1085 cctccattcttcagactttgtctactacagagaagcgccag 1125
DB 1597 CGTCCATATTCGATCTTCTCCACTATGAAAGAGACAG 1557

RESULT 14
AAL08157/c
ID AAL08157 standard; cDNA; 307 BP.

XX AAL08157;
AC
XX 07-DEC-2001 (first entry)
DT
XX Human breast cancer expressed polynucleotide 614.
DE
XX Human; breast cancer; cell marker; cytostatic; ss.
KW
XX Homo sapiens.
OS
XX WO200151628-A2.
PN
XX 19-JUL-2001.
PD
XX
XX 10-JAN-2001; 2001WO-US00798.
PF
XX
XX 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
PI
XX WPI; 2001-451856/48.
DR
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
PT
XX
XX Claim 1; Page 191-192; 3695pp; English.
PS
XX
XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
XX Sequence 307 BP; 85 A; 79 C; 55 G; 88 T; 0 other;
SQ

Query Match 8.5%; Score 173.2; DB 22; Length 307;
Best Local Similarity 73.8%; Pred. No. 3.2e-35;
Matches 220; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 734 gcacagacgcgcgaacagcgctggacagcgcttcctcgatgagcgaaggggc 793
DB 305 GTACTGAGCATGGTAATAAGCGTTTGGATGACGCTTACCGTTCCCTGAATGGGAAGGCC 246
QY 794 ccgtctacctctcctcagcgtcaatggagtgaggcattttgtgggtggccagatga 853
DB 245 CACTATATTTACTCTCTAGTGTGAATGGCAGTGGGACATTTTGTGGAGTGGCTGAATGA 186
QY 854 agtcccccgctgactacgycaccagtgccggggctggtctcagacagaagtggaaggga 913
DB 185 AGTCTGTTGTGAGCATATTAATCGGTATGCTGGTGTCTGCTCAGGATTAAGTGAAGGCA 126
QY 914 agttgatgtccagtggtatttttgaaggatgtaccacaataaccagtcocggcacatca 973
DB 125 AATTGAAAGTTAAATGGATCTTTGTCAAAGATGTTTCCCAATAACCAATTACGGCATATTC 66
QY 974 ggcctggaataacgacacaaacccgggtcacaaaactccgggacacccagggagtgcc 1031
DB 65 GCTTAGAAAAATATGACAAACAAACGGTTTACCAATTCAAGGGACACTCAAGAGGTACC 8

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 14:57:24 ; Search time 1621.09 seconds
(without alignments)
16884.827 Million cell updates/sec

Title: US-09-877-633-2

Perfect score: 2028

Sequence: 1 caaaggacaagataataa.....agtcacagacgtgtgatgga 2028

Scoring table: IDENTITY NUC

Capop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	674.6	33.3	1109	10 BE894361	BE894361 601432947
2	650.8	32.1	852	10 BE892982	BE892982 601434159
3	648.8	32.0	1014	9 AV079675	AV079675 AV079675
C 4	644.6	31.8	661	9 AL574115	AL574115 AL574115
C 5	633.4	31.2	1068	9 AL515908	AL515908 AL515908
C 6	628.8	31.0	684	9 AW001012	AW001012 wr91e06.x
7	626.6	30.9	932	9 AV726565	AV726565 AV726565
8	623.8	30.8	1479	10 BM423045	BM423045 PLATEL D1
C 9	601.8	29.7	664	10 BE646471	BE646471 7e87d05.x
10	594.6	29.3	689	10 BG911605	BG911605 602812729
C 11	583.2	28.8	600	9 AW387914	AW387914 MR4-ST011
12	581.8	28.7	808	10 BE892829	BE892829 601433961
13	580.4	28.6	736	10 BT091481	BT091481 602858960
14	571	28.2	669	9 AW964022	AW964022 EST376095
C 15	555	27.4	582	10 BT262492	BT262492 602953767
16	553.4	27.3	929	10 BG249197	BG249197 602361623
C 17	551	27.2	563	9 AW387761	AW387761 MR4-ST011

ALIGNMENTS

RESULT 1

BE894361
LOCUS 601432947F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3918154 5', mRNA linear EST 20-OCT-2000
DEFINITION BE894361 mRNA sequence.
ACCESSION BE894361
VERSION BE894361.1 GI:10356650
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1109)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9745 row: n column: 11
High quality sequence stop: 725.
Location/Qualifiers
1. 1109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3918154"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 291 a 303 c 273 g 242 t

ORIGIN

Query Match 33.3%; Score 674.6; DB 10; Length 1109;
Best Local Similarity 92.9%; Pred. No. 4e-124;
Matches 763; Conservative 0; Mismatches 49; Indels 9; Gaps 5;

QY 552 aaacgtccagcctaattctgccccagctcgaatccccaccccgctctgaaacactgaa 611
DB 1 AAACGTCACAGCCTAATCTGCCCCAGCGTCGAATCCCACCCGCTCTTGAACAACTGAA 60

QY 612 ggcgtctacagctacaacccccgaaagatttgagtggaatctgaaagggcggtgtgtt 671
DB 61 GGCCTGCTCAGAGTACAGCCGAAAGAGTTTGAGTGAATCTGAAAGCGGGCGTGT 120

QY 672 catcatcaagagctactctgagagacacatccaccgc-tccattagtagtccatctggt 730
DB 121 CATCATCAAGAGCTACTCTGAGGACGACATCCACCCGCTCCATTAAGTAGTCTCCATCTGGT 180

QY 731 gtacacagagcagcgcaacaaagccttgacagcgcttccgctgcatgagcagcaag 790
DB 181 GTAGCAGAGCAGCAGCAGCAGCGCTTGACAGCGCTTCCGCTGCATGAGCAGCAAGC 240

QY 791 ggcctgtctacbtctcttccagcgtcaatgggagtggtggtatcttctggtggcgagga 850
DB 241 GGCCTGCTACCTGCTCTCTCAGCGTCAATGGGAGTGGCAATTTTGTGGGGTGGCCGAGA 300

QY 851 tgaagtcctcccggtgagactcagcagcagtcagcggtctgctcaggaagtggaag 910
DB 301 TGAAGTCCCGCGTGGAGTACGGCAGCAGTCCCGGGTCTGGTCTCAGGACAACTGGAAGG 360

QY 911 ggaagttgatgtccagtgagatttttgaagatgtacccaataaccagctccgcaca 970
DB 361 GGAAGTTTGATGTCCAGTGGATTTTGTAAAGATGTACCCAATACCAGTCCGGCACA 420

QY 971 tcaagtgagagataacagacaacaccggtccacaactcccgagacccagaggtgc 1030
DB 421 TCAGGCTGGAGATAACGACAACAAACCGGTCAAACTCCCGGGACACCCAGGAGGTGC 480

QY 1031 ccttgaagaaagcgaagtgctgaataattatcagttcctcacaagcacacaactcca 1090
DB 481 CCTTGAAGAAAGCCAGCAAGTCTGTAATAATTATCAGTTCTCTACAAGCACACAAGTCCA 540

QY 1091 tcttcagcagcttctcactacagaaagcgca-gaggagggaggtggtgcgaag 1149
DB 541 TCTTCAGCAGCTTTGCTCACTACGAGAGCGCCAGGAGGAGGAGGTGTCGCCAAGG 600

QY 1150 aacgcagagtcgaacaaa-caatgagggcgaaacagtttcttacctgtcttaacagttt 1208
DB 601 AACGCAGAGTCGAACAAACCAACCAATGAGGCGCAACCAAGTATCTTACATGCTTCAAGCTTT 660

QY 1209 gactttga--aaacagtttaaacacagtgctgtgtcagctccagtggtcgtccgltg 1266
DB 661 GACTTTGAACAACAGTTTAAACCAACAGTGTGCTTGTGTCAGTCCAGTGTGTCGCCGTG 720

QY 1267 cgggggtgagtggtgcattcttgccttcttctgtcttg---attttgcccagatgga 1322
DB 721 CGGGGTTGAGTCTTGTCTTTGGCCTTTTCATGCTCCGTTGAATTTTGGCCCGCATGGG 780

QY 1323 tctgcatttatttgtagtcttcttctctgtattataatactctgt 1363
DB 781 TCTGCCCTTCAATGGAGGTTCTCCACAGGCATATAACCCGT 821

RESULT 2

BE92982 BE92982 852 bp mRNA linear EST 20-Oct-2000
LOCUS 601434159F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919118 5',
DEFINITION mRNA sequence.
ACCESSION BE92982
VERSION BE92982.1 GI:10353885
KEYWORDS EST.
•SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 852)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC/DCTP/DTP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LIAW5748 row: f column: 15

High quality sequence stop: 643.

Location/Qualifiers

FEATURES

source
1..852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3919118"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: Noti;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 223 a 199 c 167 g 263 t
ORIGIN

Query Match 32.1%; Score 650.8; DB 10; Length 852;
Best Local Similarity 94.9%; Pred. No. 2.2e-119;
Matches 748; Conservative 0; Mismatches 32; Indels 8; Gaps 7;

QY 969 catcagcctggagataaacgacacaaacccggtcacaaactcccgagaccccgaggg 1028
DB 1 CATCAGCCTGCAGANTAAACGACACAAACCCGTCACAAACTCCCGGACACCCAGGAGGT 60

QY 1029 gcccttagaaaaagcgaagcaagtgctgaaaaattatcagttcctcacaagcacaaacctc 1088
DB 61 GCCCTTAGAAAAAGCCAAAGCAAGTGTGTAATAATTATCAGTTCTTACAAGCACACAACCTC 120

QY 1089 catcttcgacgaacttgctcactacagagaagcgcca-gaggagagagagtggtgcgcaa 1147
DB 121 CATCTTCGAGGACTTTGCTCCTACAGAAAGCCGACGAGGAGGAGGTGTGCGCAA 180

QY 1148 ggaacggcagagtcgaaacaaacaaatgagggcgaaacacagttcttaccttctaaagctt 1207
DB 181 GGAACGGCAGAGTCGAAACAACAACAAACAGGCGCAACAGTTTCTTACATGTTTAAAGCTT 240

QY 1208 tgactttgaaaacagtttaaaaacacagtgcttggttcagctccagtggtgcctccgtgc 1267
DB 241 TGACTTTGAAAACAGTTTAAACACAGCTGCTTGGTCAGCTCCAGTGTGTCGCCGTGC 300

QY 1268 ggggggtgagtggtgcattcttgccttcttgcgttgatttttgcacagatggatctgc 1327
DB 301 GGGGGTTGAGTGTGCAATCTTTCCTTTCGTTGATTTTGGCCACAGATGGATGTGC 360

QY 1328 atttattgtacttttcttctatgtattataatcctgtagaagtcacataaagaagagatt 1387
DB 361 ATTTATTGTACTTTTCTATGTATTATAATCCTGTAGAAGTCACTAATAAAGAGGAT-T 419

QY 1388 ttttttgcagcttatcaatcacagactgacttaagtgaagttaagtataccttaaaac 1447
DB 420 TTTTGTGTGTCAGCTTATCAATCAGACTGATCTAATGTGAATGTAAAGTATCTCTTAAAAAC 479

QY 1448 aaagcatctattttggcagaaaattggttcttaaatccagtcatttgatat-tctgtgag 1506
DB 480 AAAGCATCTATTTTGGCAGAAATTTGTGTTCTTAAATTCAGTCAATTTGATATCTCTGTGAG 539

QY 1507 acttcattttctcattccttttattgcttttttagcaaacataaagaaccatgagtcattt 1566
 |||||
 Db 540 ACTTCATATTTCTCATCCCTTATTGCTTTTATAGCAACATAGAAACCATGAGTCATT 599
 |||||
 QY 1567 tgcatttagagttattcgtataaatactctctgaaataactaaatacaaaaggttaagt 1626
 |||||
 Db 600 TGTCAATTTAGAGTAGTCTGTGATAAAATCTCTTGAANTACTGAAATCAAA--GGTTAATGA 657
 |||||
 QY 1627 ttttcttcattctgatttgcattttattattctgtctatcgtgtctaa-agtgcctaattta 1685
 |||||
 Db 658 TTTCTGTCATTCATGATTTGCTCTCTTAATCATCTGCAATCGGGCTAAACAGTGTCTACTTTA 717
 |||||
 QY 1686 -ccattttagtttttctgctagacagataaacttttaattttcaaatgttgcagacatt 1744
 |||||
 Db 718 CCCCATTCGACTTCTCTCTAGACCGCTTACCCCTT-ATTTCCACATTCGCGCGACCTTT 776
 |||||
 QY 1745 tttttttt 1752
 |||||
 Db 777 TCTTCTCT 784
 |||||

RESULT 3
 AV079675
 LOCUS AV079675 1014 bp mRNA linear EST 24-OCT-2001
 DEFINITION AV079675 Mus musculus stomach C57BL/6J adult Mus musculus cDNA
 clone 2210410K23, mRNA sequence.
 ACCESSION AV079675
 VERSION AV079675.2 GI:16381140
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1014)
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koye,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)
 On Jun 25, 1999 this sequence version replaced gi:5211123.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome, 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES
 source

Location/Qualifiers
 1..1014
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2210410K23"
 /clone_lib="Mus musculus stomach C57BL/6J adult"
 /sex="male"
 /tissue_type="stomach"
 /dev_stage="adult"

BASE COUNT 266 a 238 c 252 g 256 t 2 others
 ORIGIN

Query Match 32.0%; Score 648.8; DB 9; Length 1014;
 Best Local Similarity 82.2%; Pred. No. 5,4e-119;
 Matches 822; Conservative 0; Mismatches 159; Indels 19; Gaps 6;

QY 406 cctctccagacagcccccagcttggctcaaccgcagctatcagagccctcagagcca 465

Db 19 COTCTCCCTGTNCAGCCCCNACCTGTGTGTC-CAGCCACAGTATCAGAGCCCTCAGCAGCCA 77

QY 466 cccagagccctgggtggccacgcaacaaagcggcgtttggcagagcagaggg 525

Db 78 CTTCAACCCCGCTGT-TGGCTCTCGAAACAGAAATGCGAGATGTGCGCAGAGTGGAGGG 136

QY 526 gctggcagcagatagcaactctctggaaacgtccagcctaattctgccccagcgtcgaa 585

Db 137 GCCAAGTGTGACAGTAACCTGTGTGGAAATGCCAACCTTCTTCTGCCCGAGTGTGAGAA 196

QY 586 tccaccccgctcttgaaaaactgaggctgctcagctacacacccgaaaagatttgag 645

Db 197 TCCCAACCTGTCTGTGAGAAACTGAAAGCTGCCACAGCTATAACCTTAAGAGTTCGAC 256

QY 646 lggaatctgaaagcggcgtgtctcatcatcagagctactclgagacacacccac 705

Db 257 TGGAAATCTTAAGAGTGGGGGGGTGTTCATCATCAAGAGCTATTCGTGAGGACACATCCAC 316

QY 706 cgtccatttaagtactccatctgtgttagcacagagcagcggaacaagcgctggacgc 765

Db 317 CGCTCCATCAAGTACTCCATCTGGTGTAGTAGTGAACACGGCAACAAAGCGCTGGACGGC 376

QY 766 gcttccgctgcattagacagaaggggcccgctctaccctgctcttcagcgtcaatggagct 825

Db 337 GCCTTCCTCCCTCCATCAGCAGCAAGGGGCGCTGTTTATCTCTCTTCAGTGTCAATGGAGGT 436

QY 826 gggcatttttgggggtggcagagatgaagtcctccgctgactacgacacccagtcgcgg 885

Db 437 GGACATTTCTGTGGGGTGGCAGAGATGAAGTCCCTGTGGACTATGGCACCACCGCTGGG 496

QY 886 gtctggtctcaggacaagtgggaagggaagtttgatgtccatggatgttattttttaaaggat 945

Db 497 GTCTGGTCTCAGCAACAAGTGGGAAGGAAAAGTTTGATGTGAAGTGGATTTTGTGAAGGAT 556

QY 946 gtacccaataaccagctccggcacatcagctggagagataaacgacaacaaacccggtcaca 1005

Db 557 GTGCCCAACAACACAGCTCGGCACATCACTGAGACTGGAGAAATTAACGACAACAACTGTGCACA 616

QY 1006 aactcccgagacacccagaggtgcctcttagaaaaagccaagcaagtcgtgaaattatc 1065

Db 617 AACTCCGTGTATACAGAGAGGTGCCCTTAGAAAAAGAAAACAACTGCTGAAGATTATC 676

QY 1066 agttcctcaagcacacaaacctccalcttccgacgactttgtcactacgagaagcgcca - 1124


```
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library is not normalized, but is the control for
the normalized libraries. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 382 a 217 c 187 g 275 t 7 others
ORIGIN

Query Match 31.28; Score 633.4; DB 9; Length 1068;
Best Local Similarity 98.09; Pred. No. 6.3e-116;
Matches 646; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

Qy 1370 cactaataaagagatttttttgcagcttcatcaatcagactgactaattgtgaaat 1429
Db 1068 CACTAATAAAGRGTA--TTTTTGTGTCAGCTTATCAATCAGATGATCTAATGTGAAT 1011

Qy 1430 gtaagtatccttaaaacaaagcatctattttggcagaaattgtgttctttaaattcagtc 1489
Db 1010 GTGAGTATCCTTAAACAAAGCATCTWTTTGGCAGGAATGTGTCTTAATTAATGCTC 951

Qy 1490 attgataattctgagacttcaatttctcatcccttattgcttttttagcaaacataa 1549
Db 950 ATTGGAWATCTGTGAGACTTTCATATTTCTCATCCCTTTATTGCTTTTGTAGCAACATAA 891

Qy 1550 gaaccatgaacttttgcattacagtaattctgataaattcttggaaataactgaa 1609
Db 890 GAACCATGAGTCATTTTGTGTCATTTAGAGTATCTGTGATAAATCTCTTGAATAACTGAA 831

Qy 1610 atcaaaagggttaaggattttttgttcattctgatttgcatttttatctgttctcgtt 1669
Db 830 ATCAAAAGGTTAATGATTTTGTGTCATCTGATTTGTCATTTTATTTATCTGTTATCGT 771

Qy 1670 ctaagtgctaatattaccatttgaattttctgtctagacagataaacttttaattttcaa 1729
Db 770 CTAAGTGCTAATTAACCATTTTATGATTTTCTGTCTAAACAGACATAACTTTTAATTTTCAA 711

Qy 1730 atttggcagacacttttttttttttgaattcttccagatctgttccact 1789
Db 710 ATTTGGCAGACACACTTTTTTTTTTTTGAANAATCTTTCCCTCCAGATCTGTGCCCCACT 651

Qy 1790 gaacagccacccttcctcactgtctcgtgtgtccgattgggctggatgggtgttggggcat 1849
Db 650 GAACAGCCACCCTCCCTTAACCTGTCTGCTCGGATTTGGGCTGATGGTGTGTTGGGCGAT 591

Qy 1850 gatgtgtgagaactgaaagtctttagtctcgttgcagggcgggcgatcttctgttg 1909
Db 590 GATGTGTGGAGGAACGTGAAGGTGCTTTAGGTCTGTTTTCAGGGTGGGGCATCTCTTTGTG 531

Qy 1910 ttgcacatctttttaaattttacaacctttcttaagaattctaatgcgclcttaagttt 1969
Db 530 TTTCACATCTTTTAAATTTTACACCTTTTCTTAAGAATTTCTAATGCCGCTTAAGTTT 471

Qy 1970 ttataccaataatgotgagctttaagttagtgatctgttagtacagacagtgtagtga 2028
Db 470 TTATACCAATAATGCTGAGCTTTAAGTGTAGGATCTGTGTAGAACACAGACAGTGTGATGCA 412

RESULT 6
LOCUS AW001012/c 684 bp mRNA linear EST 08-MAR-2000
DEFINITION wr9le06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2495074 3'
similar to TR:064526 064526 YUP812R.13 PROTEIN. ; mRNA sequence.
ACCESSION AW001012
VERSION AW001012.1 GI:5847928
KEYWORDS EST.
```

```
human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 684)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1194 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 476.
Location/Qualifiers
source
1..684
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2495074"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 132376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 175 a 178 c 160 g 171 t
ORIGIN

Query Match 31.0%; Score 628.8; DB 9; Length 684;
Best Local Similarity 97.8%; Pred. No. 5.2e-115;
Matches 669; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

Qy 725 tctggtgtagcacagcagcgcaaacagcgctggacagcgcttccgctgcatgagca 784
Db 683 TCTGTTGTAGCACAGAGCTCGCAAAACAGCGCTCGCACAGCGCTTCGGCTGCATGAGCA 624

Qy 785 gcaagggcccgctctacgtctcttcagcgtcaaatggagtgggcatttttgggggtg 844
Db 623 GCAAGGGCCCGCTCTACCTGCTCTTCAGCGTCAATGCGAGTGGGCATTTTTCGCGGTG 564

Qy 845 ccgagatgaagtcctcccgctgactacgcagcgccagtcgctgtgtctcagggacaagt 904
Db 563 CCGACATCAAGTCCCGCTGGACTATGGCAGCTGCGGGGCTCTGGTCTCAGGACAGT 504

Qy 905 ggaaggggaagtgttgatgccagtggtttttttaaagatgtacccaataaccagctcc 964
Db 503 GGAAGGGGAAGTTTGTATGTCAGTGGATTTTGTTAAGGATGTACCAATAACCAGCTCC 444

Qy 965 ggcacatcaggctggagataaacgacaacacccgggtcacaaactccggggacacccagg 1024
Db 443 GGCACATCAGGCTGGAGATAACGACAACACCGGTCAACAACTCCCGGGACACCCAGG 384

Qy 1025 aggtcccttagaaaaagcgaagtcgtgataaattatcatgttctctacagcacacaa 1084
Db 383 AGGTCCCTTAGAAAAAGCCAAAGCAAGTGTCTGAAATAATATCATGTTCTTACAAGCACACA 324

Qy 1085 cctccatc-ttcgacgactttgtctactacgagaagcgcca-gaggagaggaggtggtg 1142
Db 324
```

```

Db 323 CCTCCATCTTTGACGACATTTGCTCAGTACGAGAAACGCCAGGAGGAGGAGGTGGTG 264
Qy 1143 cgaaggaacgagagtcgaacaaacaaatgagggcggaacgaatttcttacatgttcta 1202
Db 263 CGCAAGGAACGGCAGAGTCGAACAACAATGAGGCGGAACAGATTTCTTACATGTTCTA 204
Qy 1203 acgtttgactttgaaacacagtttaaacacacgctgtgtgtgtgtgtgtgtgtgtgtgt 1262
Db 203 ACCTTTGACTTTGAAACACAGTTTAAACACACGCTGTGTGTGTGTGTGTGTGTGTGTGT 144
Qy 1263 cgtgcgggggtgagtgatgcatcttttgccttttcttcttcttcttcttcttcttctt 1322
Db 143 CGTGCAGGAGGTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 84
Qy 1323 tctgcattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1382
Db 83 TCTGCATTTATTTGTAATTTTCTATGTATTAATATCCCTGTAGAACTCACTAATAAGGA 24
Qy 1383 gtattttttttgtcagcttatca 1406
Db 23 GTA-TTTTGTGTCAGCTTATCA 1

RESULT 7
AV726565 932 bp mRNA linear EST 17-OCT-2000
LOCUS AV726565 HTC Homo sapiens cDNA clone HTCAxE08 5', mRNA sequence.
DEFINITION AV726565
ACCESSION AV726565
VERSION AV726565.1 GI:10835986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Gu,X., Peng,Y., Song,H., Huang,Q., Yang,X., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Shen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..932
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCAxE08"
/clone_lib="Htc"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 240 a 166 c 179 g 340 t 7 others
ORIGIN
Query Match 30.9%; Score 626.6; DB 9; Length 932;
Best Local Similarity 97.0%; Pred. No. 1.4e-114;
Matches 679; Conservative 0; Mismatches 17; Indels 4; Gaps 4;

Qy 1329 ttatttgcatttttctgtatttataacccctgtagaagtcactataaagagatttt 1388
Db 2 TTTACTCGTACTTTNCTATGATTATTAATCTGTAGAGTCACTAATAAGGAGTA-TT 60
Qy 1389 tttttgtcagcttatcaatcagactgatcctaagtgtgaatgtgaatgtccttaaaaaa 1448

```

```

Db 61 TTTTGTGTCAGCTTATCAATCAGACTGATCTAATGTGAAATGTAAGTATCTTAAAAACA 120
Qy 1449 aagcalctatttggcagaaaattgtcttcttaaatcaagtcatttggatattcgtgagac 1508
Db 121 AACATCTATTTTGGCAGAAAATTTGTCTTAAATTCAGTCATTTGATATCTGTGAGAC 180
Qy 1509 ttcataattctcatcccttttatttgcttttttagcaacaataagaaacacatgagtcatttg 1568
Db 181 TTCATATTTCTCATCCCTTTATTTGCTTTTGTAGCAACATAAGAAACCATCAGTCATTTTG 240
Qy 1569 tcatttagagtagtctctgataaaaaatctcttgaaatctactgaaatcaaaaaggttaattgatt 1628
Db 241 TCATTTAGAGTATCTCGATAAAAAATCTCTGAAAATACTGAAATCAAAAGGTTAATGATTT 300
Qy 1629 ttgttcattctctgatttgcattttatttattcttcttcttcttcttcttcttcttcttctt 1688
Db 301 TTTGTTCAATTTCTGATTTGTCATTTTATCTGTTATCTGTTATCTGTTATCTGTTATCTG 360
Qy 1689 atttgattttctctgtagacagataaacttttaatttttcaaattttggcagacactttttt 1748
Db 361 ATTTGATTTTCTGCTAGACAGATACTCTTAATTTTCAATTTTGCACACACTTTNTT 420
Qy 1749 ttttttttgaaaaatcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1808
Db 421 TTTTGTGTTTGAATAATCTTCTCTCCAGATCTGTTGCCACTGAACAGCCACCCGTCCTC 480
Qy 1809 actgctcctgggtccgactgggctggatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1868
Db 481 ACTGTCCTGGTGCCGATTTGGGCTGGATGGTGTGTGGGGCATGATGTGTGGAGAAACCTGG 540
Qy 1869 agtgctcttaggtctgtgttcagggtcgggcatctcttcttcttcttcttcttcttcttct 1928
Db 541 ACCTGCTTTAGTCTGTTTCAGGGTCGGGCATTTCTTTGGTGCTT-CACATCTTTTACA- 598
Qy 1929 ttacacaccttttcttaagaattctaatgcctgtcttaagtattttataccataatgtctgag 1988
Db 599 TTTTACACCTTTTCTAANGACATCTAATGCGCTCTTAAG-TTTTATACCAATAATGCTGAG 657
Qy 1989 ctttaagttaggtctgtgttagtacagacagtgatgtgga 2028
Db 658 CTTTAAAGTGTAACATCTGGTAGTACACAGAGTGTGTATGGA 697

RESULT 8
BM423045 1479 bp mRNA linear EST 29-JAN-2002
LOCUS PLATE1_D11 Rhesus Macaca mulatta cDNA, mRNA sequence.
DEFINITION BM423045
ACCESSION BM423045
VERSION BM423045.1 GI:18392539
KEYWORDS EST.
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
Cercopitheinae; Macaca.
REFERENCE 1 (bases 1 to 1479)
AUTHORS Katze,M.G., Bumgarner,R., Korth,M., Feldman,R., Amjadi,M. and
Holzman,T.
Expressed sequence tags from Rhesus macaque spleen
Unpublished (2002)
Contact: Holzman T
Katze Lab
University of Washington
Box 358070, Seattle, WA 98195-8070, USA
Tel: 206 732 6156
Fax: 206 732 6055
Email: ted@locke.hs.washington.edu
PLATE1_D11 1479 bases; 23.2 mean phred score.
FEATURES
Source
1..1479
Location/Qualifiers
/organism="Macaca mulatta"
/db_xref="taxon:9544"

```

AUTHORS		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov Seq primer: -40UP from Gibco High quality sequence stop: 459.		
FEATURES		Location/Qualifiers	
source	1..664		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="3292137"		
	/clone_lib="NCI-CGAP_C111"		
	/tissue_type="B-cell, chronic lymphocytic leukemia"		
	/lab_host="DH10B"		
	/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGCTTACCAATCTGAAGTGGAGCGCGCATGCTTTT		

FEATURES
source

BASE COUNT
ORIGIN

Query Match 29.7%; Score 601.8; DB 10; Length 664;
Best Local Similarity 98.0%; Pred. No. 1.2e-109;
Matches 641; Conservative 0; Mismatches 8; Indels 5; Gaps 3;

```
QY 1220 cagtttaaacacgtgtgttgcagctccagtgctgtccctgctgggggggttgagtg 1279
|||||
Db 184 CAGTTTAAACACGCTGTGCTGTGCTCAGCTCCAGTGTGCTGCCGTCCGGGGTTCAGTG 125

QY 1280 ttgatcttttgcctttcttgcgttgatttttggccagatgatctgcatatttattgtac 1339
|||||
Db 124 TTGCATCTTTGCTTCTTGTGCTGCTGATTTTGGCCAGATGTCATGTCATTTATTGTAC 65

QY 1340 tttttctatgtattataacctgttagaagtcactaataaaggagattttttt 1393
|||||
Db 64 TTTTCTATGATTAATAATCTGTGAGAGTCACATAATAAAGAGATATTTTTTT 11

RESULT 10
LOCUS BG911605 689 bp mRNA linear EST 05-JUN-2001
DEFINITION 502812729P1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944846
5', mRNA sequence.
ACCESSION BG911605
VERSION BG911605.1 GI:14292081
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10890 row: 1 column: 07
High quality sequence stop: 663.
FEATURES
Location/Qualifiers
source 1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4944846"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q
loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 190 a 154 c 157 g 188 t
ORIGIN
Query Match 29.3%; Score 594.6; DB 10; Length 689;
Best Local Similarity 97.0%; Pred. No. 3.3e-108;
Matches 670; Conservative 0; Mismatches 14; Indels 7; Gaps 6;

QY 870 cgggaccagtgccggggtctgtctcaggacaagtggaagggaagtgtgatgtccagtg 929
|||||
Db 1 CGGCACCACTGCGGGGTCTGCTCTCAGGACAAGTGGGAAGGGAAGTTTGTATGTCACGTG 60

QY 930 gattttgttaaggatgtaccacaataaccagctccggcgcacatcagctgagataacga 989
|||||
Db 61 GATTTTGTTAAGGATGTACCCAAATAACCCAGCTCCGGGCACATCAGGCTCGAGAATAACGA 120

QY 990 caacaaaccggtcacaaactccgggacaccgacgaggtgctcttagaaaaagccaagca 1049
|||||
Db 121 CAACAAACCGGTACAAACTCCCGGACACCCAGGAGGTGCCCTTAGAAAAAGCCCAAGCA 180
```

```
QY 1050 agtgtgaaataattacagttctctacaagcacacaacctccatcttgcagcagctttgctca 1109
|||||
Db 181 ACTGCTGAAATTTATCAGTTTCTTACAAGCACACAACCTCCATCTTCGACGACTTTCTCTCA 240

QY 1110 ctacagagaagcgcacca-gaggaaggagggtggtgccaagggaacggcagagtcgaaacaa 1168
|||||
Db 241 CTACGAGAAGCCCAAGGAGGAGGAGGTGGTGGCCAAGGAACGGCAGAGTGGAAACAA 300

QY 1169 acaatgaggcgcaaccagttctctacatgtctctaaagtttgcactttgaaacagtttaaa 1228
|||||
Db 301 ACAATGAGGGCGCAACCAAGCTTCTTACATGTCTTAACAGTTTGACTTTGAAAAACAGTTTAAA 350

QY 1229 acaagtgtcttggtcagctccagtgctgctcccgctgcggggggttgagtggttgcactct 1288
|||||
Db 361 ACACGCTGTCTTGGTCAGCTCCAGTGTGTCGTCGCCGTGCGGGGTTGAGTGTGTGCATCTT 420

QY 1289 tgcctttctgtctgtgatttttggccagatggatctgcatlttatttgcacttttctat 1348
|||||
Db 421 TGCCCTTCTTGTGCTGTGATTTTGGCCAGATGATCTGCATTTATTTGTACTTTTCTAT 480

QY 1349 gtattataatcctgtagaagtcactaataaaggagatttttttgcagcttataaat 1408
|||||
Db 481 GTATTATAATCTGTAGAAGTCACATAAAGAGGTA--TCTCCCGGTGAGCTTATCAAT 538

QY 1409 cagactgatctaatgtgaatgttaagtatccttaaaacaaagcatctat-tttggcaga 1467
|||||
Db 539 CAGACTGATCTAATGTGACATGTAAAGTATCTTAAAAACAAGCATCTATCCTCGGCAGA 598

QY 1468 aattgtgtcttaaatcagtcatttga-tattctgtgagacttcatat-ttctcatccc 1525
|||||
Db 599 AATTGTGTTCTTAAATTCAGTCATCCGACTATTCTGTGACACTTCATATGCTCTCATGCC 658

QY 1526 ttatttg-ctttttgacaaacataagaagacc 1555
|||||
Db 659 TTTATTTCCTTTTACCAACATTAGAAAGC 689

RESULT 11
LOCUS AW387914 600 bp mRNA linear EST 04-PEB-2000
DEFINITION MR4-ST0119-071099-010-F01 ST0119 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW387914
VERSION AW387914.1 GI:6892573
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 600)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-ST0119-
071099-010-F01&t3=1999-10-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 581.
Location/Qualifiers
source 1..600
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0119"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
```

Site_2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 124 a 169 c 151 g 156 t
ORIGIN

```
Query Match      28.8%; Score 583.2; DB 9; Length 600;
Best Local Similarity 99.3%; Pred. No. 6.1e-106;
Matches 596; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 676 atcaagagctactctgagcagcatccacccgctccatcaagtaactcctcatctggtgagc 735
Db 600 ATCAAGAGCTACTCTGAGGACGACATCCACCGCTCCATTAACTACTCCATCTGGTGAGC 541

Qy 736 acagagcaacgcaacagcgcctcgagcgccttcgcgtcatgagcagcaagggccc 795
Db 540 ACAGAGCAGCGCATCAAGCGCTGGACAGCGCTTCGGCTGCATGAGCAGCAAGGGGCC 481

Qy 796 gtctacctgctcttcagcgtcaatggagtggtgcatcttttgggggtggccgagatgaag 855
Db 480 GTCTACCTGCTCTTACGCGTCAATGGGAGTGGGCATTTTCTGGGGTGGCCGAGATGAAG 421

Qy 856 tcccccggtgactcggcaccagtgccggggtctggtctcagcagcaagtggaagggaag 915
Db 420 TCCCGCGTGGACTACGGCACCAGTGC CGGGGTCTGGTCTCAGGACAACTGGAAGGGGAG 361

Qy 916 ttgatgtccagtggtatttttgaagatgtaccataaccagctccgcgcacatcagg 975
Db 360 TTTGATGTCCAGTGGATTTTGTGTAAGGATGTACCCAATAACCAGCTCCGGGTACATCAGG 301

Qy 976 ctgagagataacgacaacacccgggtcacaaactcccggggacacccagaggtgccctta 1035
Db 300 CTGAGAGATAACGACAACAAACCCGGTCAACAACTCCCGGGACACCCAGGAGTGCCTTTA 241

Qy 1036 gaaaaagcgaagtgctgaaatattatcagttctctcaagcacaacacctccatcttc 1095
Db 240 GAAAAAGCAAGCAAGTCTCAAAATATATCATGTTTCTTACAAGCACACAACTCCATCTTC 181

Qy 1096 gacgactttgctactacgagaagcgcca-gaggagagggaggtggtgcgaaggaacgg 1154
Db 180 GACGACTTTGTCTACTACGAGAGCGCCAGGAGGAGGAGGTGTTGTCGCGCAAGGAACGG 121

Qy 1155 cagagtcgaacaaacaaatgagggcgcaaccagtttttcaatgttctaacttgacttt 1214
Db 120 CAGAGTCGAAACAAACAATGAGGGCGAACCAGTTTCTTACATGTTCTTAACGTTTGACTTT 61

Qy 1215 gaaaacagtttaaaacagtgcttggtcagctccagtgctgctcccgctgcggggggtt 1274
Db 60 GAAACACAGTTTAAACACAGTGTGCTTGGTCAAGTCCAGTGTGCTGCCGTGCGGGGTT 1
```

```
RESULT 12
BE892829 601433961F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919191 5',
LOCUS mRNA sequence.
DEFINITION BE892829.1 GI:10353393
ACCESSION BE892829
VERSION BE892829.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
```

Tissue Procurement: ATCC/DCTD/DFP
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9748 row: i column: 16
High quality sequence stop: 648.

FEATURES
source Location/Qualifiers
1..808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3919191"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SpOR6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 222 a 172 c 163 g 251 t
ORIGIN

```
Query Match      28.7%; Score 581.8; DB 10; Length 808;
Best Local Similarity 88.8%; Pred. No. 1.2e-105;
Matches 711; Conservative 0; Mismatches 77; Indels 13; Gaps 7;

Qy 977 tggagataacgacaacaaacgggtcacaaactccgggacacccagagaggtgccttag 1036
Db 10 TGGAGATAACGACAACAAACCGGTCAAACTCCCGGACACCCAGGAGGTGCCCTTAG 69

Qy 1037 aaaaagccaagtgctgaaattatcagttctctacaagcacacacctccatcttcg 1096
Db 70 AAAAAGCCAAAGTGTGTAATATTATCAGTCTCTAGACGACAACTCCATCTTCG 129

Qy 1097 acgactttgctcactcagcagcagcgcca-gaggagagggaggtggtgcgaaggaacggc 1155
Db 130 ACGACTTTGCTCATTACGAGAGCGCCAGGAGGAGGAGGTGTCGCAAGGAACGGC 189

Qy 1156 agagtcgaacaaacaaatgagggcgcaaccagttttctacatgttctaactgttgactttg 1215
Db 190 AGAGTCGAACAAACAAATGAGGGCGAACAGTTTCTTACATGTTCTTAACGTTTGACTTG 249

Qy 1216 aaaaacagtttaaaacagcgtgcttggtcagctcagctgctcccgctgcggggggtg 1275
Db 250 AAAACAGTTTAAACACAGTGTGCTTGGTCAAGTCTCCAGTGTGTCGTCGCGGGGGTGG 309

Qy 1276 agtggtgcatctttgctcttctgtcggtgattttggccagatggatctgcatattatt 1335
Db 310 AGTGTTGCATCTTTGCCCTTCTTGTGCTGTGATTTTGGCCAGATGGATCTGCATTTATTT 369

Qy 1336 gtaacttttctatgtattataatcctcgttagaagtcactaataaaggagattttttttg 1395
Db 370 GTACTTTTCTATGTATTATATATCTGTAGAAGTCACTAAATAAGGAGTA-TTTTTTTTG 428

Qy 1396 tcagcttatcaatcagactgacttaagtgaataatgaagtatccttaaaaaacaaagcatc 1455
Db 429 TCAGCTTATCAATCAGACTGATCTAATGTGAATGTAAAGTATCCTTAAACAAACAAAGCATC 488

Qy 1456 tatttggcagaataatgtgtctttaaattcagctcatttgataattctgtgagacttcatat 1515
Db 489 TATTTGGCAGAAATGTGTTCTTAANTCAGTCATTTGATATTTCTGTGAGACTTCATA- 547
```

```
Qy 1516 ttctcatcccttattgttttttagcaacataaagaacacatga-----gtcattttgtc 1570
Db 548 TTCTCATCCCTTATTGCTTTTGTAGCAAAACATAAAGAAACCAATGAGCTCATTTGGTC 607

Qy 1571 atttagagtattctgatataa-aatctcttgaaataactgaaatacaaaaggttaaatgatttt 1629
Db 608 ATTTAGAGTATTCGTGATAACAACTCTCTTGAAATACTGGAACTCCACAGGGTTATGATTTT 667
```



```
Best Local Similarity 97.4%; Pred. No. 1.6e-103;
Matches 591; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1396 tcaacttataatcagactgatcctaagtgaatgtgaatgtatctcttaaaacaagcattc 1455
Db 1 TCACGCTTATCAATCAGACTGATCTAATGTGAATGTGAATGTATCTCTTAAACAAGGATC 60

QY 1456 tatcttgccagaaattgttctttaaattcaagtcatttgatattctctgagacttcatt 1515
Db 61 TATTTTGGCAGAAATGTGTTCTTAAATTCAGTCAATTGTGATATTCTGTGAGACTTCATAT 120

QY 1516 ttctcatccctttattctcttttttagcaaaacaataagaacaacatgagtcattttgtcattta 1575
Db 121 TTCATATCCCTTTATTCCTTTTATAGCAACATAAGAACCATGATGATCTTTTGTCAATTA 180

QY 1576 gagtattctgataaaatctcttgaataactgtaactgaaatcaaaagggttaattgatttttttctc 1635
Db 181 GAGTATTCTGTAAANAATCTCTTGAAANAATCTGAAATCAAAAGGTTAATGATTTTTTGTTC 240

QY 1636 attctgatttgcattttattattctgttattcgttctaaagtgtctaattttaccatttgcatt 1695
Db 241 ATTCTGATTTGTCATTTTATTATCTGTTATCGGTCCTAAAGTGCTAATTTACCCATTTGAT 300

QY 1696 ttctctgctgagacagataaactttttaaatttttcaaatgttgagacac-tttttttttttt 1754
Db 301 TTTTCTCTAGACAGATAACTTTTAAATTTTCAAATTTGGCAGACACTTTTTTTTTTTTTT 360

QY 1755 ttgaaatctttcttccagatctgttcccactgtaacagcaaccgctccctcacttc 1814
Db 361 TTTGAAANAATCTTCTCTCCAGNCTGTGGCCACTGAACAGCCACCCGTCCTCAGTGC 420

QY 1815 ctggtgtccgattgggtggtgatggtgttggggcatgagtgtgaggaactggaaggtgc 1874
Db 421 CTGTGTCCGATTTGGCTGGATGTGTGTGGGCATGATGTGTGGAGAACTGGAAGGTGC 480

QY 1875 tttaggtctggttcagggctcgggcatctcttctgttctgacatctcttttaaatcttaca 1934
Db 481 TTTAGGCTGTGGTTCAGGGTGGGGCATACTTTGTCGTCGGCACATCTTTTAAATTTTACA 540

QY 1935 cctttcttaagaattctaatgcccgtcttaagtttttatcccaataatgctgagctttaa 1994
Db 541 CCTGTTTCAAGAATCTTAATGCAAGTTTATGCTGTTTATACCAATAATGCTGAGCTTTTA 600

QY 1995 gttgagg 2001
Db 601 ATGGTGG 607

RESULT 15
BI262492/c
LOCUS 602953767T1 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5087992 3',
DEFINITION mRNA sequence.
ACCESSION BI262492
VERSION BI262492.1 GI:14822752
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 582)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rstrasb@femail.nih.gov
Tissue Procurement: Dr. Louis Staudt, M.D., Ph.D. cDNA Library
Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

Search completed: July 29, 2002, 17:18:42
Job time: 8478 sec

```
Plate: LLCMI846 row: e column: 17
High quality sequence start: 27
High quality sequence stop: 580.
FEATURES
source
1. 582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5087992"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; site_1: XhoI; site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(C). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 142 a 161 c 133 g 146 t
ORIGIN
Query Match 27.4%; Score 555; DB 10; Length 582;
Best Local Similarity 99.0%; Pred. No. 2.5e-100;
Matches 569; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 779 tgagcagcaaggggcccgctctacgtctcttcagcgtcaatggagtggtggtttttgtg 838
Db 582 TGAGCAGCAAGGGCCCGCTTACCTGCTCTTTCAGCGTCAATGGGAGTGGGCAATTTTGTG 523
QY 839 gggTggccgagataaagtcctccctgactgacgacacagccaggtctggctcagg 898
Db 522 GGGTGGCCGAGATGAAGTCCCGCTGGACTACGGACCAAGTCCCGGGCTCTGGTCTCAGG 463
QY 899 acaagtggaaagggaagtttgatgtcccagtggtttttttaaaggatgacccaataacc 958
Db 462 ACAAGTGGAGGGCAAGTTTCATGTCACATGCGATTTTGTAAAGGATGATACCCAATAACC 403
QY 959 agctccggccacatcaggtggagataaacgacaacacccggtcacaaactcccgggaca 1018
Db 402 AGCTCCCGCACATCAGGCTGGAGAAATAACGACAACAAACCCGGTCCACAACCTCCCGGACA 343
QY 1019 cccaggagggtgccccttagaaaaagcccaagcagtgctgaaattatcagttcctcacagc 1078
Db 342 CCCAGGAGGTGCCCTTAGAAAAAGCCAAAGCAAGTCTGTAATAATATCATGTTCTCTCAAGC 283
QY 1079 acacaacctccatcttcgacgaactttgctcactacgagaagcgcca-gaggagaggagg 1137
Db 282 ACACAACCTCCATCTTCGACGACTTTGCTCACTACGAGAGCGCCAGGAGGAGGAGG 223
QY 1138 tggTgcgaaggaaacgcagaggtcgaaacaacaataatgaggggcgaaccagttcttacaatg 1197
Db 222 TGGTGCAGCAAGGAACGGCAGAGTCTGAACAACAATAAGGCGCAACCACTTCTTACATG 163
QY 1198 ttctaaagtttgactttgaaacacagttttaaacaacagtgctgctcagctccagtgctg 1257
Db 162 TTCTAAGCTTTGACTTTTGAAAAACAGTTTTAAACACAGTGTGCTTGTGTAGCTCCAGTGTGT 103
QY 1258 cgtcccgctggggggttgagtggttgcatctttgctcttctctgtgtgatttttgccag 1317
Db 102 GGTCCCGTGGGGGTTGAGTGTGTGATCTTTGCTTTTCTTGTGCTGTTGTTTGTCCAG 43
QY 1318 atggaatctgcaatttatttgggtactttttctatgtat 1352
Db 42 ATGATCTGCATTTATTGTGTAATTTCTCGGATGTAT 8
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2002, 15:15:39 ; Search time 56.49 Seconds
(without alignments)
8818.282 Million cell updates/sec

Title: us-09-877-633-2
Perfect score: 2028
Sequence: 1 caaaaggacaagataataaa.....agtcagacagtgatgga 2028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2028	100.0	2028	2	US-08-933-750C-75
2	2028	100.0	2028	3	US-08-234-613-75
3	135.4	6.7	2675	1	US-08-232-079-1
c 4	53.8	2.7	19124	2	US-08-487-826B-13
5	51	2.5	7218	1	US-08-232-463-14
c 6	50.6	2.5	6243	2	US-08-056-075-1
7	45.6	2.2	615	4	US-08-998-416-186
8	45.2	2.2	658	4	US-08-998-416-595
9	45.2	2.2	1620	2	US-08-461-775-10
10	45.2	2.2	1620	3	US-09-031-606-10
11	45.2	2.2	2668	2	US-08-461-775-11
12	45.2	2.2	2668	3	US-09-031-606-11
c 13	45	2.2	660	1	US-07-991-867B-32
c 14	45	2.2	660	1	US-08-107-755A-32
c 15	45	2.2	660	2	US-08-544-332-32
c 16	45	2.2	1511	1	US-07-991-867B-8
c 17	45	2.2	1511	1	US-08-107-755A-8
c 18	45	2.2	1511	2	US-08-544-332-8
c 19	45	2.2	4810	3	US-08-852-629-11
c 20	45	2.2	4838	3	US-08-852-629-15
c 21	44.8	2.2	6124	4	US-08-213-419B-3
c 22	43.4	2.1	7218	1	US-08-232-463-14
23	42.4	2.1	636	4	US-08-998-416-1137
24	42.4	2.1	837	4	US-08-998-416-288
c 25	42.4	2.1	51952	3	US-08-947-823-1
26	42.2	2.1	19124	2	US-08-487-826B-13
27	42	2.1	2576	1	US-08-471-033-35

28	42	2.1	2576	2	US-08-471-044-35	Sequence 35, Appl
29	42	2.1	2576	2	US-08-463-483A-35	Sequence 35, Appl
30	42	2.1	2576	2	US-08-471-046A-35	Sequence 35, Appl
31	42	2.1	2576	2	US-08-470-566B-35	Sequence 35, Appl
32	42	2.1	2576	2	US-08-469-334-35	Sequence 35, Appl
33	42	2.1	2576	3	US-09-300-529-35	Sequence 35, Appl
34	42	2.1	2655	1	US-08-471-033-17	Sequence 17, Appl
35	42	2.1	2655	2	US-08-471-033-26	Sequence 26, Appl
36	42	2.1	2655	2	US-08-471-044-17	Sequence 17, Appl
37	42	2.1	2655	2	US-08-471-044-26	Sequence 26, Appl
38	42	2.1	2655	2	US-08-463-483A-17	Sequence 17, Appl
39	42	2.1	2655	2	US-08-463-483A-26	Sequence 26, Appl
40	42	2.1	2655	2	US-08-471-046A-17	Sequence 17, Appl
41	42	2.1	2655	2	US-08-471-046A-26	Sequence 26, Appl
42	42	2.1	2655	2	US-08-470-566B-17	Sequence 17, Appl
43	42	2.1	2655	2	US-08-470-566B-26	Sequence 26, Appl
44	42	2.1	2655	2	US-08-469-334-17	Sequence 17, Appl
45	42	2.1	2655	2	US-08-469-334-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-933-750C-75
; Sequence 75, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933.750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2028 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNOBNOT03

CLONE: 1573677
US-08-933-750C-75

Query Match		100.0%	Score 2028:	DB 2:	Length 2028:		
Best Local Similarity		100.0%	Pred. No. 0:				
Matches 2028:		Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:		
QY	1	caaaagacaagaataaagtacaaaatgggttcgtttacatcagaaaggatcagtttcag	60				
DB	1	CAAAAGGACAGATATAAGTACAAATGGTTCGTACATCAGAAAGATACAGTTCATC	60				
QY	61	acnaatgactttgagccctaccctacttgacagtcacaatcagagtaacagttaccctcaa	120				
DB	61	ACAANTGACTTTGAGCCCTACCTTACTTGGACAGTCAAAATCAGAGTAACAGTTACCCTCAA	120				
QY	121	tgaggaacccctaccttcacagctattacccgcggtcccatggatttcctactcctca	180				
DB	121	TGAGGACCCCTACCTTGTCCAGCTATTACCCGCGTCCCATTTGGATTTTCCTTACTCCCTCA	180				
QY	181	atgaggtccgtgggtacttgaggggacccctccgatttccatcactccacacacagac	240				
DB	181	ATGAGGCTCCGTGGTCTACTGAGGGGACCCCTCCGATTCCATACCTCACCACTACGGAC	240				
QY	241	agctcagtaacggagaccatcattttatgcagatgctgtttttggcgacctgggggccc	300				
DB	241	AGCTCAGTAAGCGAGACCATCATTTTTATGCACGATGCTGTTTTGGGCAGCCTGGGGGCC	300				
QY	301	tgagggaacacatctacagcacagttcaattttttcccttgaaaacctggcttcag	360				
DB	301	TGGGAACACATCTATCAGCACAGGTTCAATTTTTTCCCTTGAAACCCCTGGCTTCTCAG	360				
QY	361	catgggggacaagtgggtctcaaggtcagcagaccagagctcagctctccccagcacag	420				
DB	361	CATGGGGACAAGTGGGTCTCAAGGTCACGACACCAGAGCTCAGGCTCTCCCGACACAG	420				
QY	421	ccccagcttgggtcnaaccgagtalacagagccctcagcagcccccagaccgctgg	480				
DB	421	CCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAGACGCCACCCAGACCCGCTGG	480				
QY	481	gttgcccccagacaagaacggcggttggcgagcgagggcggtggcgagcgatagc	540				
DB	481	GTTGCCCCACGANCAAGAACGGCGGCTTGGCGAGCGGAGGGCTGGCAGCGATAGC	540				
QY	541	aactctcctggaaacgtccagcttaattctgccccagcgctcgaaatcccaaccgcctctt	600				
DB	541	AACTCTCCTGGAAACGTCACGCTAATTCTGCCCCAGCGTCGAATCCCAACCCGCTT	600				
QY	601	gaaaaactgaaggtctcacagctacaacccgaaagagtttgagtggaatctgaaaaagc	660				
DB	601	GAAAACTGAAGGCTGCTCACAGCTACAACCCGAAAGAGTTTGAGTGGAAATCTGAAAAAGC	660				
QY	661	gggcgtgtgtctctatcaagagctactctgagggagacatccaccgctcccatcaagtac	720				
DB	661	GGCGGTGTGTTTCATCATCAAGAGCTACTCTGAGGAGCAGCATCCACCCGCTCCATTAAGTAC	720				
QY	721	tccatctgtgtgtagcacagacggcgaacaagcgcttgagcagcgcttccgctgcatg	780				
DB	721	TCCATCTGTGTAGCACAGACAGCGCAACAAGCGCCTGGACAGCGCCTTCCGCTGCATG	780				
QY	781	agcagcaaggggcccgtctacactgctctcagcgtaaatgggagtgggcatctttgtggg	840				
DB	781	AGCAGCAAGGGGCCGCTCTACCTGCTCTTACGCGTCAATGGGAGTGGCATTTTTCGTGGG	840				
QY	841	gtggccgagatgaagtcccccgtggactacggcagcagtcgcgggggtctgggtctcaggac	900				
DB	841	GTGGCCGAGATGAAGTCCCCCGTGGACTACGGCACAGCTGCCGGGGTCTGGTCTCAGGAC	900				
QY	901	aagtggaaggggaagtttgtatgtccagtggaattttttgtaaggaatgtacccaataaccag	960				
DB	901	AAGTGAAGGGGAAGTTTGATGTCCAGTGGATTTTTTTGTTAAGGATGTACCCAATAACCAG	960				
QY	961	ctccgggacacatcaggctgggagaataacagacaacaccgggtcacaaactccggggacacc	1020				

RESULT 2

DB	961	CTCCGGCACATCAGGCTGGAGAATAACGACAACAAACCGTCAAAACTCCCGGACACC	1020
QY	1021	caggaggtgcccttagaaaaagccaagcaagtgtctgaaaaattatcagttcctcaagcac	1080
DB	1021	CAGGAGTGCCTTTAGAAAAAGCCAAAGTGTCTGAAAAATTATCA3TTCCTTACAAGCAC	1080
QY	1081	acacactccatcttcacagacttgcctcactacgagaagccagagagagagagaggtgg	1140
DB	1081	ACAACTCCATCTTCAGACACTTTGCTCACTACGAGAAGCGCCAGAGAGGAGGTGG	1140
QY	1141	tgcgcaaggaacggcgagagtcgaacaacaaatgaggggggaacacagttcttatcagtttc	1200
DB	1141	TGCGCAAGGAACGGCAGAGTCGAACAACAAATGAGGGCGGAACAGTTTCTTACATGTTTC	1200
QY	1201	taacgtttgactttgaaacacagtttaaaacacagctgtgcttgggtcagctccagtgctcgt	1260
DB	1201	TAAAGTTTGACTTTGAAACAGGTTTAAACACGTTGCTTGGTCAGTCCAGTGTGTGCT	1260
QY	1261	cccgctggggggttgagtggtgcatcttgcctttctctgtctgtgaatttttggccagatg	1320
DB	1261	CCCGTGGGGGTTGAGTGTGTCATCTTTGCCCTTTCTTGTGCTGATTTTGTGCCAGATG	1320
QY	1321	gactgcattattctgaacttttctatgtattataatcctgttagaagtcacataaag	1380
DB	1321	GATCTCATTTATTGTTACTTTTCTATGTATATATCTCTGTAGAAAGTCACTAATAAAG	1380
QY	1381	gaactatttttttgcagcttcaatcagactgacttaagtgaatgaatgaatgaatcct	1440
DB	1381	GAGTATTTTTTTTGTGAGCTTATCAATCAGACTGATCTAANTGTGAANTTAAGTATCCT	1440
QY	1441	taaaaaaagcatctattttggcgagaaatgtgtcttctaaatctcattctattgatctc	1500
DB	1441	TAAAAACAAGCATCTATTTTGGCACAAATGTGTCTTAAATTCAGTCAFTTGTATATTC	1500
QY	1501	tgtagacttcaatctcactcccttattgcttttttttagcaacaataagaaccatgag	1560
DB	1501	TGTGAGACTTCATATTCTCATCCCTTATTGCTTTTGTAGCAACAATAAGAACCATGAG	1560
QY	1561	tcaatttgcatttagagttattctgataaatactctgaaaaataactgaaaaatacaaggtt	1620
DB	1561	TCAATTTGTCAATTTAGAGTATTCTGATAAANTCTCTTGAAAAATCTGAAATCAAAAGGTT	1620
QY	1621	aatgatttttgcattctgatttgcatttttattattattctgttctggtctaaagtgccta	1680
DB	1621	AATGATTTTGTTCATCTGATTTGTCATTTTATTATCTGTTATCGTCTAAAGTGTCTA	1680
QY	1681	atttaccacttgaattttctctgcttagacagataaactttttaaatttttcaaatcttggcagac	1740
DB	1681	ATTTACCCATTTTGATTTTCTGCTAGACAGATAACTTTTAAATTTTCAATTTTGGCAGAC	1740
QY	1741	actttttttttttttgaaaaatcttctccagatctgttgcacctgaaacagcagacc	1800
DB	1741	ACTTTTTTTTTTTTTTGGAAAAATCTTTCCTTCAGATCTGTTGCCCTGAAACAGCCACC	1800
QY	1801	cgctccactgctcctgggtgcagttggggtggatgggtgttgggggagatggtggag	1860
DB	1801	CGTCCCTCACTGCTCCTGGTGTCCGATTGGGCTGGATGGTGTGGGGCATGATGTGTGGAG	1860
QY	1861	gaactggaaggtgctttaggtctgggttcaggggtcggggcattcttctgtgtttcacatct	1920
DB	1861	GAACTGGAAAGGTGCTTTAGGTCTGGTTCAGGGTGGGGCATTTCTTGTGTTGCATCT	1920
QY	1921	tttttaattttacacacttttcttaagaaattctaatgccctcttaagtcttttataccata	1980
DB	1921	TTTTTAATTTTACACCTTTTCTTAAAGAAATCTAANTGCCGCTCTTAAGTTTATATCCANTTA	1980
QY	1981	atgctgagctttaagtgtaggatctggttagtacagacagtgatgagga 2028	
DB	1981	ATGCTGAGCTTTAAGTGTAGGATCTGTTAGTACACAGACAGTGTGATGGA 2028	

US-09-234-613-75
; Sequence 75, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2028 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNOBOT03
; CLONE: 1573677
US-09-234-613-75

Query Match 100.0%; Score 2028; DB 3; Length 2028;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2028; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 caaaggcaagataataaagtacaaaatggttcgtttacatcagaaggatcagttcatg 60
Db 1 CAAAGGCAAGATAATAAGTACAAAATGGTTCTTACATCAGAAGGATACAGTTTCATG 60
Qy 61 acaatgactttgagccctacccttactggaagtcacaaatcagaagtaacagttaccctcaa 120
Db 61 ACAATGACTTTGAGCCCTACCTTACTGGACAGTCAAAATCAGAGTAACAGTTACCCCTCAA 120
Qy 121 tgagagccctaccctgctccagctattaccgcccgcctcattggatttccctactccctca 180
Db 121 TGAGAGCCCTACCCTGCTCCAGCTATTACCCGCCGCTCCATTGGATTTCCTTACTCCCTCA 180
Qy 181 atgaggtccgtggtctactcagaggaccctccgattccatcacctaccctacggac 240
Db 181 ATGAGGTCCGTGGTCTACTGCAGGGGACCCCTCCGATTTCATACCTACCACTACGGAC 240

Qy 241 agctcagtaacggagaccatcatctttatgcagatgctgtttttgggcagcctgggggcc 300
Db 241 AGCTCAGTAACGGAGACCATCATTTTATGCAGATGCTGTTTTTGGCAGCCTGGGGGCC 300
Qy 301 tggggaacaacatctatcagcacaggttcaatttttccctgaaaaacccctggttctcag 360
Db 301 TGGGGAACAACATCTATCAGCACAGGTTCAATTTTTTCCCTGAAAAACCCCTGGTTCTCAG 360
Qy 361 catggggacaagtgggtctcaagctcagcacagccagagctcagcctctccacacacag 420
Db 361 CATGGGGACAAGTGGGTCTCAAGGTCAGCAGACCAGAGCTCAGCCTCTCCAGCACAG 420
Qy 421 cccccagctttggctcaaccgcagtatcagagccctcagcagcccccagaccgcgtgg 480
Db 421 CCCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAGCACGCCACCCAGACCCGCTGG 480
Qy 481 gtlgcccacacagcaaaaacggcggtttggcgagagcgagggtgagcgagataagc 540
Db 481 GTTGCCCCACGCAACAGAAACGCGCGTTTGGCGAGAGCGAGGGCTTGGCAGCGATAGC 540
Qy 541 aactcctggaacgctccagcctaattctgccccagcgctcgaatcccccctgctcct 600
Db 541 AACTCTCTGGAAACGTCAGCCTAATTTCTGCCCCAGCGTCGAATCCACCCCGCTCCTT 600
Qy 601 gaaaaactgaagctgctcacagctcacaccccgaaagagtttgagtggaatctgaaaaagc 660
Db 601 GAAAAACTGAAGCTGCTCACAGCTACAAACCCGAAAGAGTTTGAAGTGAATCTGAAAAAGC 660
Qy 661 gggcggtgttcatcatcaagagctactctgaggagcacatcacgcgtccatttaagtac 720
Db 661 GGGCGGTGTTCATCATCAAGAGCTACTCTGAGGACGACATCCACCGCTCCCATTAAGTAC 720
Qy 721 tccatctggttagcacagagcacggaacagccctgagcagcctctccgctgcagc 780
Db 721 TCCATCTGGTGTAGCACAGAGCACGGAACAAGCGCTGGACAGGCCCTTCCCGTGCATG 780
Qy 781 agcagcaaggggcccgctcactctcttcagctcgaatgggagtggggcatttttgtggg 840
Db 781 AGCAGCAAGGGGCCCGCTCTACCTGCTCTTCAGCGCTCAATGGGAGTGGGCATTTTTGTGGG 840
Qy 841 gtgcccagagatgaagtccccctggactacggcaccagtgcccggggtctggtctcaggac 900
Db 841 GTGGCCGAGATGAAGTCCCCCGTGGACTACGGCACAGTGCCGGGTCTGTGCTCTCAGSAC 900
Qy 901 aagtgaagggaagtgttgatgccagtgattttttaaagatgatacccaataaccag 960
Db 901 AAGTGAAGGGGAAGTGTGATGTCAGTGGATTTTTTTTAAGATGTATCCCAATAACCCAG 960
Qy 961 ctccggcacatcaggtgagaaataacgacaacaaccggtcacaaaactcccgggacacc 1020
Db 961 CTCGGCACATCAGGCTGGAGAATAAGGACAACAACCGGTACAAACTCCCGGGACACC 1020
Qy 1021 cagggggtgccccttagaaaaagccaagcgaagtgcgtgaaaattatcagttccctacaagac 1080
Db 1021 CAGGAGGTGCCCTTAGAAAAAGCCAAGCAAGTGTGAAATTAATCATAGTTCTCTACAAGCAC 1080
Qy 1081 acaactccatcttcagcagactttgctcactacgagaagcgcagagggagggaggtgg 1140
Db 1081 ACAACTCCATCTTCAGCAGACTTTGCTACTAGAGAAGCGCCAGAGGAGGAGGAGGTGG 1140
Qy 1141 tgcgcaagaaacggcagagctcgaacaaacaatgagggcgaaaccagttcttaccatgttc 1200
Db 1141 TCGCAAGAAACGGCAGAGTTCGAAACAAACAATAGGGCGCAACCACTTTCTTACATGTTC 1200
Qy 1201 taactgttgactttgaaaaacagtttaaacacagtgctgtgttcaggtccagtggtcgt 1260
Db 1201 TAACGTGTGACTTTGAAAAACAGTTTAAAAACAGTGTGCTTGGTTCAGCTCCAGTGTCTGT 1260
Qy 1261 cccgtgcgggggttagtggttgcatttgcctttcttcttcttcttcttcttcttccagagc 1320
Db 1261 CCCGTGCGGGGTTAGAGTGTTCATCTTTGCCCTTCTTGTGCTGATTTTTTGTCCCAAGATG 1320
Qy 1321 gatcgcattttattgtactttttctattattataactctgtagaagtcactaataaag 1380


```
Db 1209 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1268
QY 1806 ctcaactgctgctgcgcattggcgtgatggtgtgggcatgatgctgtgagagact 1865
Db 1269 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1328
QY 1866 gaaagtgcttgatgctgctgctgaggtcggtggcattctgtgtgtgcacatctttta 1925
Db 1329 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1388
QY 1926 aatttacaccctttcttaagaattcotaatgcgcgtcttaagtcttttatccaa 1978
Db 1389 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAA 1441

RESULT 6
US-09-056-075-1/c
: Sequence 1, Application US/09056075
: Patent No. 5955368
: GENERAL INFORMATION:
: APPLICANT: Johnson, Eric A.
: APPLICANT: Bradshaw, Marite
: APPLICANT: Rood, Julian
: TITLE OF INVENTION: Expression System for Clostridium
: NUMBER OF INVENTIONS: Species
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles & Brady
: STREET: 1 South Pinckney Street
: CITY: Madison
: STATE: WI
: COUNTRY: US
: ZIP: 53701-2113
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/056,075
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J.
: REGISTRATION NUMBER: 27386
: REFERENCE/DOCKET NUMBER: 960296.95238
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6243 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 3770..4013
: OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
: OTHER INFORMATION: plasmid RP4"
US-09-056-075-1

Query Match 2.5%; Score 50.6; DB 2; Length 6243;
Best Local Similarity 47.6%; Pred. No. 0.0011;
Matches 149; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 1459 ttggcagaaattgttcttaattcagtcatttgatattctgtgagacttcatttc 1518
Db 1506 TATGAAAAAATGGCGCTTTATTAAGGCTCATTTTATATCTTTCTTCAAGATTA 1447
QY 1519 tcatccctttattgctgttttttagcaaacataagaaccatgagtcattttgtcatttagag 1578
```

```
Db 1446 TATAATATAAAAAATTTTTCACAACTTTAAATAAAAAATATTTTATATATTTTATT 1387
QY 1579 tattctgataaatactctctgaaataactgaaatcaaaaggttaaatgatttttgcatt 1638
Db 1386 TTTTATATTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATT 1327
QY 1639 ctgattgctcaatttattatctgattatcggtctaaagtgctaaattaccacattgatttt 1698
Db 1326 TATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATT 1267
QY 1699 tctgctagacagataacttttataatttttcaaatcttgccagacacattttttttttttg 1758
Db 1266 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATT 1207
QY 1759 aaaaatcttccctt 1771
Db 1206 TTATTTTATTATT 1194

RESULT 7
US-08-998-416-186
: Sequence 186, Application US/08998416
: Patent No. 6239284
: GENERAL INFORMATION:
: APPLICANT: Philippssen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Knechtle, Philipp
: APPLICANT: Rebeschung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Mgigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 186:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 615 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1074RP
US-08-998-416-186
```

```
Query Match 2.2%; Score 45.6; DB 4; Length 615;
Best Local Similarity 46.0%; Pred. No. 0.0075;
Matches 197; Conservative 0; Mismatches 224; Indels 7; Gaps 1;

Qy 1330 ttatttgacattttcttgatgattataatcctgtagaagtcactaataaaaggagatattt 1389
Db 84 TTATTATCATTTATTAATAATTAATTAATTTGATTATTAATCACTATTAATAAATTAATA 143
Qy 1390 tttttgcagcttcaatcagactgactaagtgaagtgaagtcactccttaaaaaaa 1449
Db 144 TATAATTTACATTAATTCATCATTAATTAATTAATTAATTAATAAATAATTAATA 203
Qy 1450 agcactattttggcagaaattgttcttaaatcagtcatttgatattctgtgagact 1509
Db 204 TGAATACATATTAGTCTATGTTCAAAATTTAAATTTAGTTATTAATAATTT-----ATT 256
Qy 1510 tcaattctcctccttttctgttcttttagcaaacataagaaccatgagtcatttgt 1569
Db 257 AGATATTAATTTTCTTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 316
Qy 1570 catttagagattcttgataaaactctcttgaaaactgaaactcaaaagggttaagtattt 1629
Db 317 TATTAATGTTTATTAATAAATAATATATTTTATTAATTAAGATTTAATTTATTAATA 376
Qy 1630 ttgttcactctgattgttcattttattatctgttgcgtctaaagtcgaatttaccoca 1689
Db 377 TTGTAATTAATTAATTTTATTAATAATCTATTTTATTAATAATTAATTAATTTGATTATAT 436
Qy 1690 ttgtattttctgctagacagataacttttaattttcctaaatttgcgcagacactttttt 1749
Db 437 ATTTAACTTTTATTAAGAAATTAATTAATTAATAATTAATTAATTAATTTTCTTATTAATA 496
Qy 1750 tttttttt 1757
Db 497 ATTTTAT 504

RESULT 8
US-08-998-416-595
; Sequence 595, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998.416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
```

```
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 595:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1408RP
US-08-998-416-595

Query Match 2.2%; Score 45.2; DB 4; Length 658;
Best Local Similarity 46.1%; Pred. No. 0.0099;
Matches 187; Conservative 0; Mismatches 218; Indels 1; Gaps 1;

Qy 1329 ttatttgacttttctcatgtattataatcctgtagaagtcactaataaaaggagattt 1388
Db 249 TTTATTATTTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 308
Qy 1389 tttttgtcagcttcaatcagactgactcctaatgtgaagtcactccttaaaaaa 1448
Db 309 TTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTGT 368
Qy 1449 aagcactattttggcagaaattgttcttaaatcagtcatttgatattctgtgagac 1508
Db 369 TGTTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 428
Qy 1509 tt-catatctcctccttttctgttcttttagcaaacataagaaccatgagtcattt 1567
Db 429 TTACTTAGTATATAATATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 488
Qy 1568 gtcatttagagattctctgataaaactctcttgaaaactgctgaaactcaaaagggttaagtatt 1627
Db 489 ATTATATAGATTGTTATTTGTGACACATAATATATATGCTATTTCTATTTCTATTTATTT 548
Qy 1628 tttttcattctgattgttgcatttttattatctgttctatctgttctaaagtcgaattacc 1687
Db 549 TATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 608
Qy 1688 catttgatttttctgtagacagataacttttaatttttcaattt 1733
Db 609 TATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 654

RESULT 9
US-08-461-775-10
; Sequence 10, Application US/08461775
; Patent No. 5858773
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```


US-07-991-867B-32

Query Match 2.2%; Score 45; DB 1; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.011;
Matches 220; Conservative 0; Mismatches 250; Indels 5; Gaps 2;

Qy	1293	ttctctgtcgttgatatttttgcgcagaatcgcgattattttgtactatttttctatgtat	1352
Db	512	ttttaaattaaattttttctataaatttataaactttttattttgtattgtatcatttt	453
Qy	1353	tataactcgtagaagtcactaaataaaggagatttttttctgcagcttatcaatcaga	1412
Db	452	aactcaactatatactaatagcaaaaacattatttaagttacccaattttattttttagataatt	393
Qy	1413	ctgatctaattggaattgaatatacttaaaacaaagcatctattttggcagaaaattg	1472
Db	392	attaaattttttttaaatttacttaattatttaattaaaaattatagtcattttatattacatgat	333
Qy	1473	tgttcttaaatcgaactttgatattctgtgagacttcattttctcattctcattcttattg	1532
Db	332	tcacaattttaaaaatttctatagaattggtagatataattttactttatattgtctattttttg	273
Qy	1533	ctttttagcaacataagaacccatgagtcattttgtcatttagagtatctcgtataaat	1592
Db	272	ttataagatatatctaaatatcttatatttttttaattttgtttataaaaatttaaaatttaata	213
Qy	1593	ctcttgaaaatactgaaatcgaaggttaagtattttttgtctcaattctgatttgtcattt	1652
Db	212	atatttttaaaatt--tgaaaatatataaacttttaataatttttcggaatatatttttaaaaatat	155
Qy	1653	tattatctgttatcgtctcaagtgtaatttaccacatttgtattttctcgtcagacagat	1712
Db	154	tattatcatataatatattata--tcgaattctcttaaaattaacataatttttttttaatatat	98
Qy	1713	aactttttaatttttcaaatattggcagacaacttttttttttttttttttttggaaactcttt	1767
Db	97	taataatttaataacattatctctgtttattattatttttttttttttttttttttaaatattataatttt	43

RESULT 14
US-08-107-755A-32/c
: Sequence 32, Application US/08107755A
: Patent No. 5721352
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Guild, Michael E.
: TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: Florida
: COUNTRY: U.S.A.
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/107.755A
: FILING DATE: 19-AUG-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/827,658
: FILING DATE: 30-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/657,584
: FILING DATE: 19-FEB-1991
: ATTORNEY/AGENT INFORMATION:

; NAME: Saliwanchik, David R.
 ; REGISTRATION NUMBER: 31,794
 ; REFERENCE/DOCKET NUMBER: UF114.C2

```

; TELFAAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 32:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 660 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: unknown
;
; MOLECULE TYPE: DNA (genomic)
US-08-107-755A-32

```

Query Match 2.2%; Score 45; DB 1; Length 660;
Best Local Similarity 46.3%; Pred. NO. 0.011;
Matches 220; Conservative 0; Mismatches 250; Indels

[illegible]

```

RESULT 15
US-08-544-332-32/c
; Sequence 32, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Guirdl, Michael E.
; TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

Search completed: July 29, 2002, 18:02:58
Job time: 10039 sec

Query Match	2.2%;	Score 45;	DB 2;	Length 660;
Best Local Similarity	46.3%;	Pred. No. 0.011;		
Matches 220;	Conservative 0;	Mismatches 250;	Indels 5;	Gaps 2;
QY	1293	ttctctgctgattgatttttccacagatggatctgcacatttattttgtaactttttctcatgtat	1352	
DB	512	TTTAAATTAATTAATTTTCTATAAATTTATAATCTTTTATTGTTATTTGATTCGATATTT	453	
QY	1353	tataactctgagagtcactataaaggagatttttttttttgcagctatcaalcaga	1412	
DB	452	AACCTCAACTACTAATAGGAATAACATTTATTAAGTTACCAAAATTTATTTTATAGATATT	393	
QY	1413	ctgctctaattgaaattgaagtatcccttaaaacaagcatctattttggcagaaattg	1472	
DB	392	ATTAATTTTTTTTAAATTTACTAAATTTATAAATAATATAGTCATTTATATTACATCAT	333	
QY	1473	tgctcttaaatcagtcatttgatattctglgagacttcattttctcaccctttattg	1532	
DB	332	TCACAATTTAAAAATCTATAGAATGTGGTAGTATAATATTACTATATTGCTATTTTGT	273	
QY	1533	ctttttgacaaacataagaacacctgagtcattttgcatttagagtgatttcgtataaaat	1592	
DB	272	TTATAAGATATATCTAAATATGTTATTTTTTTAAATTTTCTATATAAATTTAAATTAATA	213	
QY	1593	ctcttgaaatactgaaatcaaaaggctaatgatttttttgctatctctgattgtcattt	1652	
DB	212	ATATTTTAAATTT--TGAATATATAAACTTTTAAATTTTCTGGAATATTATTAAAAATAT	155	
QY	1653	tattattctgtatcggctctaaagtctcaatttaccattttgatttttctgtagacagat	1712	
DB	154	TATTATCATATAATATATA--TGCAATCTCTCTAAATTAACATAATTTTTTTTAAATAT	98	
QY	1713	aacttttaatttttccaaatttggcagacacttttttttttttttttttttttttttttt	1767	
DB	97	TAAATATTAAACATTTATCTCTGTTTATATTATTTTATTTTAAATTTAAATATTTT	43	

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 15:06:14 : Search time 2520.46 Seconds
(without alignments)
16837.806 Million cell updates/sec

Title: US-09-877-633-2
Perfect score: 2028
Sequence: 1 caaagagcaagataataa.....agtcacagcagtgtgatgga 2028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	2028	100.0	2028	6	AR116045	Sequence
2	1595	78.6	2615	9	AB055518	Homo sapi
3	1595	78.6	2952	9	BC003681	Homo sapi
4	1593.4	78.6	2429	9	AK000398	Homo sapi
5	1590.8	78.4	2695	9	BC016920	Homo sapi
c	870	42.9	140152	9	HSJ963E22	Human DNA
7	549.2	27.1	168012	2	AC025565	Homo sapi
8	549.2	27.1	172451	2	AC106748	Homo sapi
c	365.8	18.0	192994	2	AC011978	Homo sapi
10	341.6	16.8	2309	9	BC002559	Homo sapi
11	341.6	16.8	2750	6	AX048109	Sequence
12	340.6	16.8	2070	9	AF155095	Homo sapi
13	340.6	16.8	2730	9	AF192968	Homo sapi
14	340.2	16.8	2079	10	BC014797	Mus muscu
c	333.4	16.4	130984	2	AC022766	Homo sapi
16	333.4	16.4	162532	2	AL645729	Homo sapi
17	328	16.2	8533	9	AK024546	Homo sapi
c	321	15.8	163627	2	AL672026	Mus muscu
c	321	15.8	194985	10	AC002406	Mouse chr
c	318.2	15.7	141372	2	AC011779	Homo sapi
21	318.2	15.7	196319	9	CNS01DT4	Human chr
22	299	14.7	1573	10	BC022932	Mus muscu
23	271.8	13.4	2907	9	AK056219	Homo sapi
c	258.4	12.7	65091	2	AC100199	Mus muscu
25	226	11.1	379	6	AX072124	Sequence
c	214.6	10.6	32785	2	AC013560	Homo sapi
27	209.6	10.3	2679	3	AY069675	Drosophil
c	195.4	9.6	113474	2	AC013932	Drosophil
29	195.4	9.6	162921	3	AC007853	Drosophil
c	195.4	9.6	181132	3	AC008206	Drosophil
c	195.4	9.6	227398	3	AE003750	Drosophil
c	169	8.3	36384	9	AC004178	Homo sapi
33	169	8.3	104423	2	HSJ824K2	Human DNA
34	169	8.3	112578	9	AL645927	Homo sapi
c	169	8.3	114868	9	HS994E9	Human DNA
37	169	8.3	129567	2	AL671618	Homo sapi
38	147.4	7.3	1840	8	AY072018	Arabidops
39	136.6	6.7	1768	8	AY056421	Arabidops
40	136.6	6.7	1836	8	AY065413	Arabidops
41	135.4	6.7	2675	6	118667	Sequence 1
c	132.8	6.5	423	6	AX330380	Sequence
43	132.4	6.5	2163	8	AF326910	Arabidops
44	125.8	6.2	2484	8	AY050425	Arabidops
c	119	5.9	101589	2	AP003897	Oryza sat

ALIGNMENTS

RESULT 1	AR116045	Sequence 75 from patent US 6132973.	2028 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	AR116045	Sequence 75 from patent US 6132973.				
DEFINITION	AR116045	Sequence 75 from patent US 6132973.				
ACCESSION	AR116045	Sequence 75 from patent US 6132973.				
VERSION	AR116045.1	GI:14096367				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 2028)					
AUTHORS	Lal, P., Hillman, J.L., Bandman, O., Shah, P., Au-Young, J., Yue, H., Guegler, K.J. and Corley, N.C.					
TITLE	Human regulatory molecules					
JOURNAL	Patent: US 6132973-A 75 17-OCT-2000;					
FEATURES	Location/Qualifiers					
source	1. .2028					
BASE COUNT	520 a 478 c 463 g 567 t					
ORIGIN						
Query Match	100.0%;	Score 2028;	DB 6;	Length 2028;		

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2028; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	caaaagcaagataataaagtacaaaatggttccttacatcagaagatcacagttcatg	60
Db	1	CAAAAGGCAAGATAAATAAAGTACAAAATGGTTCCTTACATCAGAAAGATACAGTTCATG	60
Qy	61	acaatgactttgagccctactctactgacagtcacaaacagagtaacagttaccctcaaa	120
Db	61	ACAATGACTTTGAGCCCTACCTTACTGGACAGTCAAAATCAGAGTAACAGTTACCCCTCAA	120
Qy	121	tgagcgacccctacctgtctcagctattacccgcctcacttggaattccttactcctca	180
Db	121	TGAGCGACCCCTACCTGTCTCAGCTATTACCCGCCCTGCCATTGGATTCTCTCCCTCA	180
Qy	181	atgaggtccgttggtctactcagagggaccctccgattccatccactcaccactcaggac	240
Db	181	ATGAGGCTCCGGTGTCTACTGCAGGGGACCCCTCCGATTCCATACCTCACCACTACGGAC	240
Qy	241	agctcagtaacggagaccatcatatttatgacgagtctgttttgggcagcctggggcc	300
Db	241	AGCTCAGTAACGGAGACCATCATTTATGCACGATGCTGTTTTTGGGCAGCCTGGGGCC	300
Qy	301	tggggaacaacatctatcagcacaggttcaaatttttccctgaaacacccctggtctcag	360
Db	301	TGGGGAACAACATCTATCAGCACAGGTTCAATTTTTTCCCTGAAAACCCCTGCGTCTCAG	360
Qy	361	catggggacaaagtgggtctcaagtccagcagaccagagctcagccctccacacag	420
Db	361	CATGGGGACAAGTGGGTCTCAAGGTCCAGCACAGACCCAGCTCAGCCCTCTCCACACAG	420
Qy	421	ccccagctttggctcaaccgagtcacagagccctcagcagcccccagaccgctgg	480
Db	421	CCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAGCACCCACCCAGACCCGCTGG	480
Qy	481	gtgccccacgcacagaaaccggcggttggggcagcgaggggctggcgagatagc	540
Db	481	GTTGCCCAACGCAACAGAAACCGCGGTTTGGGCAGAGCGGAGGGGCTGGCAGCGATAGC	540
Qy	541	aactctctggaaagctcagcctaatctgcgccagcgctcgaatccaccaccctcctt	600
Db	541	AACTCTCTGGAAAGCTCAGCGCTAAATCTGCCGCCAGCGCTCGAATCCACCCCGCTCTT	600
Qy	601	gaaactgaagggtgctcacagctacacaccgaaagagtttgadtgaaatctgaaaagc	660
Db	601	GAAAACTGAAGCGTCTCACAGCTACAAACCCGAAAGAGTTTGAGTGAATCTGAAAAGC	660
Qy	661	gggcgtgtgtcatcatcaagagctactctgaggcgacatccacgcgtccaatlaagtac	720
Db	661	GGGCGTGTGTTCATCATCAAGAGCTACTCTGAGGACGACATCCACCGCTCCCAATTAAGTAC	720
Qy	721	tccatctgggtgtagcacagcagcgcaacaagcgcttgacagcgccctccgctgcatg	780
Db	721	TCCATCTGGTGTAGCACAGAGCACGGCAACAAGCGCTTGACAGCGCCTTCCGCTGCATG	780
Qy	781	agcagaagggggccgctacgtctcttcagcgtcaaatgggagtgggcatttttgtagg	840
Db	781	AGCAGAAGGGGCCCGTCTACCTGTCTTCAGCGTCAATGGGAGTGGGCATTTTTTGTGGG	840
Qy	841	gtggccgagatgaagtccccctggactacggcaccagtgccggggctggtctcaggac	900
Db	841	GTGGCCGAGATGAAGTCCCGCTGACTACGGCACAGTGGCGGGGCTGTGGTCTCAGGAC	900
Qy	901	aagtggaaaggggaagtttgatgtccagtggaattttgttaaggatgtaccacaataccag	960
Db	901	AAGTGAAGGGGAAGTTTGATGTCCAGTGGATTTTTTGTAAAGATGTACCCCAATAACCCAG	960
Qy	961	ctccggcacatcaggctggagaaataacagacaaacacgggttcacaaactccccggacac	1020
Db	961	CTCCGGCACATCAGCTGGAGAAATAAGCACAAACAAACCGGTTCACAAACTCCCCGGACACC	1020
Qy	1021	caggagggtcccttagaaaaaaccaagcaagtgtctgaaaaattatcagttcctcacagacc	1080
Db	1021	CAAGGAGGTCCCTTAGAAAAACCAAGCAAGTGTCTGAAAAATTATCAGTTCTCTCAACAGCAC	1080

Db	1021	CAGGAGTGCCCTTACAAAAAGCCAAAGCAAGTGCTGAAATATTATCATCTCTCAACAGCAC	1080
Qy	1081	acaactccatcttcgagacatttgcctcactacgaaagcgcagagggagggaggtg	1140
Db	1081	ACAACCTCCATCTTCGACGACTTTGCTCACTACGAAAGCCGACAGAGAGGAGGTGG	1140
Qy	1141	tgccaaagaaagcggcagagtcgaaacaaacaaatgagggcgaaaccagttctctacatgttc	1200
Db	1141	TGCCGAAGGAACGGCAGAGTCGAAACAAACAATGAGGGGGAACCACTTCTTACATGTTTC	1200
Qy	1201	taacgttgactttgaaacacagtttaaaacacagtcgttgcttggtcagctccagtgatctc	1260
Db	1201	TAACGTTTGACTTTGAAAACAGTTTAAACACAGTGTGCTTGGTTCAGCTCCAGTGTGCTCT	1260
Qy	1261	cccgtagcgggggttagtgatcatcttggcccttctctgcttctgatttttggccagag	1320
Db	1261	CCCGTGGCGGGGTGAGTGTTCATCTTTGCCCTTCTTGTGCTTGAATTTTGGCCAGATG	1320
Qy	1321	gactgcatttatctgtacttttctatgtattataaactcctgtagaagtcactcaataaag	1380
Db	1321	GATCTGCATTTATTTGTAATTTTCTATGTAATAATCTGTAGAAAGTCACATAAAG	1380
Qy	1381	gagtatcttttctgcagcttatcaatcagactgatcacaatgtgaaatgtaagtatcct	1440
Db	1381	GAGTATTTTTTTTGTACGCTTATCAATCAGACTGATCTAAATGTGAATGTAAATGATCCT	1440
Qy	1441	taaaacaaagcatctattttggcagaaattgttcttcaaattcagtcatttgatctc	1500
Db	1441	TAAACAAAGCATCTATTTTGGCAGAAATGTGTCTTAAATTCAGTCAATTTGATATATTC	1500
Qy	1501	tgtagacttcataatttctcatcccttatttgccttttagcaaaacaagaacacatgag	1560
Db	1501	TGTGAGACTTCATATTTCTCATCCCTTTATTCGCTTTTAGCAAAACATAAGAAACCATGAG	1560
Qy	1561	tcattttgtcatcttagagatctctgatataaactctctgaaataactgaaatacaaaaggtc	1620
Db	1561	TCAATTTGTCTATTTAGAGTATTTCTGATAAAATCTCTGAAAATCTGAAATCAAAAGGTT	1620
Qy	1621	aatgatttttctcatctgatttgatcttcttcatctgttctcgtgctaaagtgcta	1680
Db	1621	AATGATTTTTTGTCTCATTTCTGATTTGTCTATTTATATCTGTTATCGGTCCTAAAGTGCTA	1680
Qy	1681	atttaaccatttgatttttctctgtagacagataaacttttcaatttttcaaatbtggcgagc	1740
Db	1681	ATTTACCACTTTGATTTTCTGCTAGACAGATAAATTTTAAATTTTCAAAATTTGGCAGAC	1740
Qy	1741	acttttttttttttttttgaaaacttttccctccagatctgttgcacactgaaacagccacc	1800
Db	1741	ACTTTTTTTTTTTTTTTTGAAAATCTTTTCTTCCAGATCTGTGTGCCACACTGAACAGCCACC	1800
Qy	1801	cgccctcactgctccctgggtgctcgattgggctggatgggtgttggggcatgagtgtggag	1860
Db	1801	CGTCCCTCACTGCTCCTGGTGTCCGATTGGGCTGGATGGTGTGGGGCATGATGTGTGGAG	1860
Qy	1861	gaactggaaaggtgctttaggtgcttcagggttcggggcattcttctgttcttgcatctc	1920
Db	1861	GAACTGGAAAGGTGCTTTAGGTCTGGTTCAGGGTCGGGCATTTCTTTGTTCTTTGCACATCT	1920
Qy	1921	tttttaaattttcaacaccttttcttaagaatactaaatgcgctcttaagttttttatcaccaata	1980
Db	1921	TTTTTAAATTTTACACCTTTTCTTAAAGAAATTTCTAAATGCCGCTCTTAAGTTTTTTATACCAATA	1980
Qy	1981	atgctgagcttttaagtgtaggactcgtgtagtcacagacagtgtaggaga	2028
Db	1981	ATGCTGAGCTTTTAAGTGTAGGATCTGGTAGTACAGACACAGTGTGATGGA	2028

RESULT 2
AB055518
LOCUS
DEFINITION
ACCESSION

AB055518
Homo sapiens
putative autoantigen-1, partial cds.
2615 bp
mRNA
linear
PRI 07-AUG-2001

```
VERSION      AB055518.1  GI:15128560
KEYWORDS
SOURCE       Homo sapiens cell_line:HeLa cell cDNA to mRNA.
ORGANISM     Homo sapiens
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2615)
AUTHORS      Onouchi,H., Muro,Y. and Tomita,Y.
TITLE        Dermatomyositis Associated with Cancer Autoantigen
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 2615)
AUTHORS      Muro,Y.
TITLE        Direct Submission
JOURNAL      Submitted (06-FEB-2001) Yoshinao Muro, Nagoya University School of
             Medicine, Department of Dermatology; Tsurumai-cho, Nagoya, Aichi
             466-8550, Japan (E-mail: ymuro@med.nagoya-u.ac.jp, Tel:81527442314,
             Fax:81527442318)
FEATURES     Location/Qualifiers
             1..2615
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /cell_line="HeLa cell"
             1..1315
             /gene="DACA-1"
             /gene="DACA-1"
             /codon_start=2
             /product="dermatomyositis associated with cancer putative
             autoantigen-1"
             /protein_id="BAB62751.1"
             /db_xref="GI:15128561"
             /translation="ARENPAFSAMGTSGSGQQTQSAYGSSYTPPSSLGGTVVDGQ
             PGFHSOTLSKAPQMNSLEOGMVLKIGDVSSAVKTVGVSVALTGVLGNGGTVV
             NMPVSKPTSWAATASPKAPQPKMKTSGPVMGGGLPPPIKHNMDLGTWDMKGPVK
             APVQAPSPQAPAPQOVAOPLPAOPPALAQPOYQSPQOPQTRWVAPRNRAAFQ
             SGAGSDNSPGNQPNSAPVESHPVLEKLAHSTNPREFWNLKSGRVFIKSTVS
             EDDIHSRISICWSTEHGKNRLDLSAPRCMSRGVYLLESVNGSHGFCGVAEMKSPV
             DYGTSAWVSQDKWKGFQVDFVYKADVPNNQLRHRLNNDKPKVTNSRDTQEVPLE
             KAKQVLKISISYKHTTIFEDDFAHYBKROBEVEVVRKERQSRNKQ"
BASE COUNT   664 a 622 c 634 g 695 t
ORIGIN
Query Match      78.6%; Score 1595; DB 9; Length 2615;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1628; Conservative 0; Mismatches 0; Indels 3; Gaps 3;
QY 400 gctcagcctctccagacagccccagcttggctcaaccgcagtgatcagagccctcag 459
DB 539 GCTCAGCCTCTCCAGCACAGCCGCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAG 598
QY 460 cagccaccagaccgctgggttgcgccacgcaacagaaacgcggtttggcagagc 519
DB 599 CAGCCACCCAGACCCGCTGGTGGCCCGCCACGCAACAGAAACGCGCGTGGGCGAGC 658
QY 520 gggagggctggcagatagcaactctctctggaaacgtccagcctaattctgccccagc 579
DB 659 GGAGGGGTGGCAGCATAGCAACTCTCTCGTGAAGGTCCAGCCTAATTCTGCCCCAGC 718
QY 580 gtcaatcccccccgctcttgaaaaacgtagagctgctcacagctacaacccccaaagag 639
DB 719 GTCAAAATCCCAACCCCGCTCTTGAAGAACTGAAGGCTGCTCACAGCTACAACCCGAAAG 778
QY 640 tttagtggaatctgaaaagcgcggtgtgttcattcatcaagagctactctgagagcac 699
DB 779 TTTGAGTGAATCTGAAAGCGGCGGTGTGTATCATCAAGAGCTACTCTGAGGACCAC 838
QY 700 atccacgcctcattaaagtactcatctgtgttagcacagagcacggcaacaagcgctg 759
DB 839 ATCCACCGCTCCATTAAGTACTCCATCTGGTGTAGCAGAGCAGCGCAACAGCGCCGTG 898
QY 760 gacagcgccttcgcgtgatgagcagaagggccgtctacctgctcttcagcgtcaat 819
|||||
```

```
DB 899 GACAGCGCCTTCCGCTGCATGAGCAGCAAGGGGCCGCTGTACTGCTCTTTCAGCGCTCAAT 958
QY 820 gggagtgggcatttttgggggtgcccagatgaagtcccccgtagctacgcagccagt 879
|||||
DB 959 GGGAGTGGGCATTTTGTGGGTGGCCGAGATGAAGTCCCGGTGGACTACGGCACCACT 1018
QY 880 gccgggggtctggctcaggacaagtgggaagggaagtlttgagtccagtggaatttttgg 939
DB 1019 GCCGGGGTCTGCTCTCAGGCAAGTGGGAAGGGAAGTTTGATGTCACGTGGATTTTTGT 1078
QY 940 aagatgtaccocaataaacacagctccggcacatcaggctgggagataaacgacaacaaacg 999
DB 1079 AAGATGTATCCCAATAAACACAGCTCCGGCACATCAAGCTGGAGAAATAGCACAACAAACCG 1138
QY 1000 gtcaaaactccccgggacacccagagtgcccttagaaaaagccaagtcgctgaaa 1059
DB 1139 GTCACAAACTCCCGGGACACCCAGGAGTGCCCTTAGAAAAAGCCCAAGTGTCTGAAA 1198
QY 1060 attacagttcctacaagcacacaacctccatctctcgacgactttgctcactacagagaag 1119
DB 1199 ATTATCAGTTCTCTACAAGCACACAACCTCCATCTCTCGACGACTTTTGTCTACCTACGAGAAG 1258
QY 1120 cggca-gaggaggaggagtggtgcgcaagaacggcgagtcgaaaaacaaatgaggg 1178
DB 1259 CGCCAGGAGGAGGAGGAGGTGGTGCAGGAAGCAAGCGCAGAGTCGAACAAACAATGAGGG 1318
QY 1179 cgaaccagttctacatgttctaactgttaacgcttgaactttgaaacagtttaaaacacgtgc 1238
DB 1319 CGAACCACTTTCTTACATGTTCTTAACGTTGACTTTGAAACACAGTTTAAACACGCTGTC 1378
QY 1239 ttggtcagctccagtgatgcgtcccggtcggggttgagtggtgcatcttgccttctct 1298
DB 1379 TTGCTCAGCTCCAGTGTCTGTCGCCGTGGGGGTGAGTGTGTCATCTTTGCCCTTCTT 1438
QY 1299 gtggtgattttggcccgatggaatcgtcatttatctgtacttttctctatgattataat 1358
DB 1439 GTCGTTGATTTTGGCCAGATGGATCTGATCTATTTATTTGTTACTTTTCTATGATTATAAT 1498
QY 1359 cctgtagaagtcaccaaataaaggagatattttttgctcagcttatcaatcagactgac 1418
DB 1499 CCTGTAGAAGTCACTAATAAAGGAGTA-TTTTGTTCAGCTTATCAATCAGACTGATC 1557
QY 1419 taatggaatgtaagtatccttaaaaaaagcatctattttggcagaaaattggttct 1478
DB 1558 TAAATGTGAATGTAAGTATCTCTTAAAAACAAGCATCTATTTTGGCAGAAATTTGTTCT 1617
QY 1479 taattcagtcatttgatctatctgtgagacttcaatttctcattcccttctatctgtttt 1538
DB 1618 TAAATTCAGTCAATTTGATATTTCTGTGAGACTTCATATTTCTCATCCCTTTATTTGCTTT 1677
QY 1539 agcaaacataagaaccatgagtcattttgcatcatttcttcattctgatttgctatttattat 1598
DB 1678 AGCAAACTATGAACACCATGAGTCATTTTGTTCATTTTGTTCATCTGATTTTATTTAT 1737
QY 1599 aaatactactgaaatcaaaagggttaagatttttttctcattctctgatttgctatttattat 1658
DB 1738 AAAATACTGAAATCAAAAGGTTAATGATTTTGTTCATCTGATTTTGTTCATTTTATTTAT 1797
QY 1659 ctgltatcgggtctaaagtgcataatttaacccattgatatttctgtctagacataaacttt 1718
DB 1798 CTGTTATCGGTCTAAAGTGTCTAAATTTTACCCATTTTCTGCTAGACATTAACCTTT 1857
QY 1719 taatttttcaaatttggcagacac-ttttttttttttttttttttttttttttttttttttttt 1777
DB 1858 TAAATTTTCAAAATTTGGCAGACACTTTTFTTTTTTTTTTTTGAATAATCTTCTCTCCAGAT 1917
QY 1778 ctgttgcccaactgaagaccacccctccctcactgtccctggtgctccgattgggctggagtg 1837
DB 1918 CTGTTGCCCACTGAACAGCCACCCCTCCCTCCTACTGCTCTGGTGTCCGATTTGGCTGATG 1977
QY 1838 gtgttggggacatgagtgtgaggaactggaaggtgctttaggtctgtgttcaggggtcgagg 1897
DB 1978 GTGTTGGGGCATGATGTGTGGAGGAACCTGGAAGGTGCTTTAGTGTCTGTGTTTCAGGGTCCGG 2037
```

```
QY 1898 cattttttgtgttgacacatctttttaaatatttccacacctttttcttaagaattcttaatgc 1957
|||||
Db 2038 CATTCCTTTGCTTCCACATCTTTTAAATTTTACACCTTTTCTTAAGAATTTCTAATGC 2097
|||||

QY 1958 cgtcttaagtatttataccaataatgctgagcttttaagtgtaggatctggtagtagcacag 2017
|||||
Db 2098 CGTCTTAAGTTTTTATACCAATAATGCTGAGCTTTAAGTGTAGGATCTGCTAGTACAGAC 2157
|||||

QY 2018 agtgtagtagga 2028
|||||
Db 2158 AGTGTGATGGA 2168
|||||

RESULT 3
BC003681
LOCUS BC003681 Homo sapiens, clone IMAGE:3453235, mRNA, partial cds. PRI 12-JUL-2001
DEFINITION BC003681
ACCESSION BC003681
VERSION BC003681.1 GI:13277545
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2952)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-ghgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 4 Row: 9 Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7020459.
FEATURES
source
1..2952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3453235"
/tissue_type="Placenta, chorioncarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
<L..1648
/codon_start=2
/product="Unknown (protein for IMAGE:3453235)"
/protein_id="AAH03681.1"
/db_xref="GI:13277546"
/translation="KGQDNKVONGSLHOKDVTVDNDFFPYLTGOSNQSNYSMSDPY
LSSYPPSGIPGPPYSLEAPWSTAGDPPIPLYTYQLSNGDHHFMDAVFGQPGGLGN
NIYQHRFFNPENAFSAWSTGSGQQQTQSSAYGSSYTPPSSILGTVVDQPGFHS
DTLSKAPQMNSLEQMGVLKIGDVSSAVKTYSVSSVALTGVLSGNGGTWNMPVS
KPTSWAATASKPAKPQPKMTKSGPVMGGGLPPPIKHNMDIGTDWNGKGPVPKAPVQ
QAPSPOAAPOPOVAOPLPAQPAALAOPOYQSPQOPOTRWVAPRNRNAAFQSGCAG
SDNSPGNVOPNSAFVESHVPLEKLAHNSYNPKFEWNLKSGRVFLIKSYSDDIH
RSIKYSIWCSTEHGKRLDLSAFRCMSKGPYLLFSVNGSHGFCGVAEMKSPVDVGTG
ACVWSQDKWKGKFDVQWIFVKDVPNNQNLRIHLENNDNKPV7NRSRDTOEVLKAKQV
```

LK11SSYKHTT5IFDDFAHYEKRRQEEVEVVRKERSRKO*

BASE COUNT 758 a 715 c 700 g 779 t
ORIGINQuery Match 78.6%; Score 1595; DB 9; Length 2952;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1628; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

```
QY 400 gctcagcctctccacacagacagcccccagcttttggtccaaacgagctatcacagccctcag 459
|||||
Db 872 GCTCAGCCTCTCCACAGCACAGCCCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAG 931
|||||
QY 460 cagccacccacagaccgctgggttgccccacgcaacagaaacgcggttttgagcagagc 519
|||||
Db 932 CAGCCACCCACAGACCCGCTGGGTGGCCCCACGCAACAGAAACCGCGGCTTTGGGCAGAGC 991
|||||
QY 520 ggggggcttgagcagatagcaactctctggaaaacgtccagcctaatctgccccccagc 579
|||||
Db 992 GGAGGGGCTGGCAGCGATAGCAACTCTCCTGGAAACGTCACAGCCTAATTTCTGCCCCAGC 1051
|||||
QY 580 gtggaatccacaccccgctcttgaaaaactgaaggctgctcacagctacaaccgaaagag 639
|||||
Db 1052 GTCGAATCCCAACCCCGCTCTTGAAAAACTGAAGCCTGCTCACAGCTACAAACCCGAAGAG 1111
|||||
QY 640 ttgagtggaatctgaaaagcgggcggtgtttcatcatcaagagctactctgaggcagc 699
|||||
Db 1112 TTTGAGTGGAACTGTAAGACGGCGGTGTGTTTCATCATCAAGAGCTACTCTGAGGAGGAC 1171
|||||
QY 700 atcacgcgtccatctaaagtactccatctggtgtagcacagagcagcgcaacagcgctg 759
|||||
Db 1172 ATCCACCGCTCCATTAAGTACTCCATCTGTTGACACAGACGACGCGCAACAGCGCGCTG 1231
|||||
QY 760 gacagcgccttcgctgcatgagcagcaaggcgcccgctctacactgctcttcacgctcaat 819
|||||
Db 1232 GACAGCGCCTTCGCTGTCATGAGCAGCAAGGGCGCGCTCTACTGCTCTTCACGGCTCAAT 1291
|||||
QY 820 gggagtgggcatcttttggtgggtggcgagatgaagtcgcccgctggactacggcaccagt 879
|||||
Db 1292 GGGAGTGGGCATTTTGTGGGGTGGCGAGATCAAGTCCCGCTGGNACTACGGCACACAGT 1351
|||||
QY 880 gccggggtctggtctcagagacaagtggaaagggaagtgtgatgccagtggaattttgtt 939
|||||
Db 1352 GCCGGGCTCTGGTCTCAGGCAAGTGCAGGGGAAGTTTCATGTCCAGTGGATTTTGT 1411
|||||
QY 940 aagatgtagcccaataaacagctccgacacatcagctgagagataacgacacaacaacccg 999
|||||
Db 1412 AAGGATGTACCCAATTAACAGCTCCGGCAGATCAGGCTGGAGATTAACGACAAACAACCCG 1471
|||||
QY 1000 gtcacaaactcccgagacacccagagggtgccccttagaaaaagccaagtcagtgaaa 1059
|||||
Db 1472 GTCACAAACTCCCGGACACCCAGAGGTGCCCTTAGAAAAAGCAAGTGTGTGAAA 1531
|||||
QY 1060 attatcagttctacaaagcacacaacctccatcttcgaagactttgctcactacgagaag 1119
|||||
Db 1532 ATTATCAGTTCTTACAAAGCACACAACCTCCATCTTCGACGACTTTGTCTACTACGAGAAG 1591
|||||
QY 1120 cgcga-gaggaggagagtggtgcgaaggcaagcagagtcgaacaacaacaatgagg 1178
|||||
Db 1592 CGCCAGGAGGAGGAGGAGGTGTGCGCAAGGAACGGCAGAGTCCGAACAACAACATGAGGG 1651
|||||
QY 1179 cgaaccagtttcttacatgttctaaagtttgactttgaaacagttttaaaccagctgtgc 1238
|||||
Db 1652 CGAACCACTTCTTACATGTTCTTACGTTTTCGACTTTGAAAAACAGTTTAAACACGTGTGC 1711
|||||
QY 1239 ttggtcagctccagtgctgttcctccgtcgcggtgggttgagtggtgcactcttgctctctt 1298
|||||
Db 1712 TTGCTCAGCTCCAGTGTGCTCCCGTGGCGGGGTGAGTGTGTCATCTTTGCTTTGCTTCT 1771
|||||
QY 1299 gtcgttgatttttgcacagatgagctgcatctatttgtaacttttctatgtattataat 1358
|||||
Db 1772 GTCGTTGATTTTTGCCAGATGGATCTGCATTTATTGTTGACTTTTCTATGTATTATAAT 1831
|||||
```


Qy	999	gggtcaaaactccgggacacccagaggtgccccttagaaaaagccaagcaagtgctgaa	1058
Db	958	GGTCACAAACTCCGGGACACCCAGGAGGTGCCCTTAGAAAAAGCCAAAGTGCTGAA	1017
Qy	1059	aattatcagttcctacaagaacacaaacctccattcttcgacgactttgctcactacagaa	1118
Db	1018	AATTATCAGTTCTACAAGCACAAACCTCCATCTTTCGACGACTTTGCTCACTACAGAA	1077
Qy	1119	gcacca-gaggaaggaagtggtgcgcaagaagaagcagagtcgaacaaacaatgaag	1177
Db	1078	GOCCAGGAGGAGGAGGTGTGCGCAAGGAGCGAGCTCGAAACAAACAATGAGG	1137
Qy	1178	gcgaacagatttctacatgttcttaacgcttgactttgaaacagtttaaaacacagtg	1237
Db	1138	CGCAACCACTTCTTACATGTTCTTAACGTTGACTTTGAAACACAGTTAAACACGTGTG	1197
Qy	1238	cttggtcagctccagtgctgcctccggtgcgggggttgagtggtgcacatttgcctttct	1297
Db	1198	CTTGGTCAGCTCAGTGTGCTGCCGTGCGGGGTGAGTGTGTCATCTTTGCCCTTCT	1257
Qy	1298	tatcggtgattttccccagatgagctgcatttatttatactttttctatgtattataa	1357
Db	1258	TGTCGTTGATTTTGGCCAGATGGAFTGCAATTTATTTGTACTTTTCTATGTATTATAA	1317
Qy	1358	tctgttagaagtcactaaataaaggagatatttttttgcagcttatcaatcagactgat	1417
Db	1318	TCTGTAGAAGTCACATAAAGGAGTA-TTTTTTTGTGCAGCTTATCAATCAGACTGAT	1376
Qy	1418	ctaattgtaagtgaatgtaacttaaaacaaagcatcatttttggcagaaattggttc	1477
Db	1377	CTAATGTGAATGTAAGTATCTCTTAAACAAAGCATCTATTTTGGCAGAAATTTGTTC	1436
Qy	1478	ttaattcagtcatttgatctctgagacttcatttccatttccatcccttttcttcttt	1537
Db	1437	TTAAATTCAGTCAATTTGATTTCTGTGAGACTTCATTTTCATCCCTTATTTGCTTTT	1496
Qy	1538	tagcaacaatagaacaccatgagtcattttgtcatttagagtattctgtataaactcttt	1597
Db	1497	TAGCAAAACATAAGAACCATGAGTCATTTTGTTCATTTAGAGATTTCTGTATAAATCTCT	1556
Qy	1598	gaataactgaatacaataaaggtaagatcttttcttctcattctgatttctcatttctta	1657
Db	-1557	GAATAACTGAATAACAAAGGTAATGATTTTGTTCATTTCTGATTTCTCAATTTTATTA	1616
Qy	1658	tctgttaccggtcctaagtgctaatttaccatttgatttttctgctagacagataactt	1717
Db	1617	TCTGTTATCGGCTAAGTGCTAATTTACCCATTTGATTTTCTGCTAGACAGATTAATTT	1676
Qy	1718	ttaatttttcaatttggcagacactttttttttttttttttttttttttttttttccat	1777
Db	1677	TTAAATTTTCAAAATTTGGCAGACACTTTTTTTTTTTTTGTGAAATCTTTTCTTCCAGAT	1736
Qy	1778	cgtgtgccactgaacagcaccccgctccctcactgctcctggtccgaatgggctggatg	1837
Db	1737	CTGTGCCCACTGAACAGCACTCGTCCCTCACTGCTCCTGGTGTCCGATTTGGCTGGATG	1796
Qy	1838	gtgttgggcagatgctgtgaggaactggaagtgactttagtctggttcagggctcgag	1897
Db	1797	GTGTTGGGCGCATGATGTGTGGAGAACTCGAAGGTGCTTTAGTGTCTGGTTCAGGGTCCGG	1856
Qy	1898	cattctttgtgtttgacatctcttttaaattttaacaccttttcttaagaattctaatgc	1957
Db	1857	CATTCCTTTCTGTTTGCACATCTTTTAAATTTTACACCTTTTCTTAAGAATTTCTAATGC	1916
Qy	1958	cgtcttaagtttttataccaataatgctgagctttaagtgtaggatactcgtgtagtacagac	2017
Db	1917	CGTCTTAAGTTTTTATACCAATTAATGCTGAGCTTTAAGTCTAGGATCTGCTAGTACAGAC	1976
Qy	2018	agtgtgtagga	2028
Db	1977	AGTGTGATGGA	1987

RESULT	5			
BC016920	2695 bp	mRNA	linear	PRI 09-NOV-2001
LOCUS	Homo sapiens, clone IMAGE:3853372, mRNA, partial cds.			
DEFINITION	BC016920			
ACCESSION	BC016920.1			
VERSION	GI:16877330			
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2695)			
TITLE	Strausberg, R.			
JOURNAL	Direct Submission			
REMARK	Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.			
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 20 Row: 0 Column: 16 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8923363. Location/Qualifiers 1..2695 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3853372" /tissue_type="Colon, adenocarcinoma" /clone_lib="NIH_MGC_65" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" <1..1403 /codon_start=3 /product="Unknown (protein for IMAGE:3853372)" /protein_id="AAH16920.1" /db_xref="GI:16877331" /translation="HSDHFMHDAVFGQGGGLGNNIYQHRFPFPENPAFSAMGTSGS QGOQTSAYGSYTPPSSLGCTVVDQPGPHSDTLKAPGMNSLEQGMVLKIGDV SSAVKTVGSVVSVALTGVLGNGGTNNMPVSKPTSWAAIASPKAPQPMKTKSG PVMGGLPPPIIKHNDGTMDNKGVPKAPYQQAAPQQAQVQVLAQPPA LAQPYQSPQPPQTRWAPNRNNAFGSGAGSDNSPGNVQNSAPSVSHPVLE KLKAHSYNPKFENLKSGRVFIISKYSEDDIHKRSIKYSIWCSTEHGKRLDSAPRC MSSKCPVLLFSVNGSGHCGVAEMKSPVDYCTAGVMSQDWKAKFQWILFVKDVP NNQLRHRLNNDNRPVNSRDTQEVPLEKAKQVLKLISSYKHTTSIFDDFAHYEKRO EEEEVYRKQRQSRNQ"			
CDS	674	a	644	c 654 g 723 t
BASE COUNT	674 a 644 c 654 g 723 t			
ORIGIN				
Query Match	78.4%; Score 1590.8; DB 9; Length 2695;			
Best Local Similarity	99.7%; Pred. No. 0;			
Matches 1625; Conservative	0; Mismatches 2; Indels 3; Gaps 3;			
Qy	400	gctcagcctctccacagacagccccagctttgctcaaccagctatcagagccctcag	459	
Db	627	GCTCAGCCTCTCCACGACAGACGCCGCCAGCTTTGCTCAACCGCAGTATCAGAGCCCTCAG	686	
Qy	460	cagccacccacagaccgcgtgggttgccccacgcaacagaaacgcggcgtttgggcagagc	519	

```
Db 687 CAGCACCCAGACCGCGTGGTTCGCCACGCAACAAACGGCGGTTTGGCAGAGC 746
Qy 520 ggaggggctgagcagatagcaactctctctggaacgctccagcctaaactctgccccagc 579
Db 747 GGAGGGCTGCAGGATAGCAACTCTCTCTGGAACGCTCCAGCCTAATCTGCCCCAGC 806
Qy 580 gtcgaatcccccctctcttgaaaactgaagctgctcacagctacaacccaaagag 639
Db 807 CTCGAATCCACCCCGCTCTTGAACAACTGAAGCTGCTACAGCTACAACCCGAAGAG 866
Qy 640 tttagtggaatcgaagaagggcggtgttctcatcatcaagagactactcgaagacag 599
Db 867 TTTGAGTGAATCTGAAAGCGGCGGTGTTCATCAATCAAGAGCTATTCTGAGGATGAC 926
Qy 700 atccacgcctcattaaagtactcactctggtgtgacagagcgcggaacaagcgcctg 759
Db 927 ATCCACCGCTCCATTAACTACTCTCATCTGCTGTAGCAGCAGCAGCGCAACAGCGCTG 986
Qy 760 gacagcgcttcctcgctgcatgagcagcaaggggcccgtctacctgctcttcagcgtaat 819
Db 987 GACAGCGCTTCCGCTGCATGAGCAGCAAGGGCGCGCTTACCTGCTCTTCAGCGTCAAT 1046
Qy 820 gggagtgggcaatttttctggtggcgagatgaagtcccccgctggactcagggcaacagt 879
Db 1047 GGGAGTGGGCAATTTTGTGGGGTGGCGGAGATGAAGTCCCGCGTGGACTACGGCACCACT 1106
Qy 880 gccgggctggtctcaggacaagtgaaggggaagtttgatgtccagtgatattttt 939
Db 1107 GCCGGGCTGCTCTCAGGACAGTGGGAAGGGGAAGTTGATGTCCAGTGTGATTTTGT 1166
Qy 940 aagatgtacccaataaccagctccgcacatcaggtgggagataacgacaacaaaccg 999
Db 1167 AAGATGTACCCAATAACCAGCTCCGGCACATCAGGCTGGAGATAACGACAACAAACCG 1226
Qy 1000 gtcacaaactccgggacaccacggaggtgcccctagaaaaagccaagcaagtcgtgaaa 1059
Db 1227 CTCACAACTCCCGGGACACCCAGAGGTGCCCTTAGAAAAAGCAACAAAGTGTGAAA 1286
Qy 1060 attacagttctcaagcacacaacctccatctctcgacgacttctgctcactacgagag 1119
Db 1287 ATTATCAGTTCCTACAAACGACACAACTCCATCTCTCGACGACTTGTCTCATTACGAGAG 1346
Qy 1120 cgcca-gaggaggagggtggtgogcaaggaaagcgcagagctcgaacaaacaatgaggg 1178
Db 1347 CGCCAGGAGGAGGAGGAGTGTGCGCAAGCAAGCGCAGAGTCGAACAAACAAATGAGGG 1406
Qy 1179 cgaaccagttcttacaatgttctaacgtttgactttgaaacagttttaaacaagctgtgc 1238
Db 1407 CGAACCCAGTTTCTTACATGTCTTAACTGCTTGAAGTTTGAACAGTTTAAACACAGTGTGC 1466
Qy 1239 ttggtcagctccagtggtgctccgctgaggtgaggtgaggttgcatctttgctctttt 1298
Db 1467 TTGTCACGCTCCAGTCTGCTGCCGCTCGGGGGTTGAGTGTGGATCTTGGCTTTCTT 1526
Qy 1299 gtcgttgattttgcccagatgagatcgtcattttattgtaactttttctatgtattataat 1358
Db 1527 GTCGTTGATTTTGGCCAGATGATCTGCATTTATTGTACTTTTCTATCTATTATTAAT 1586
Qy 1359 cctgtagaagtcactaaataaggagatatttttttgcagcttatcaatcagactgac 1418
Db 1587 CCTGTAGAAGTCACTAAATAAAGGAGTA-TTTTGTGTGTCAGCTTATCAATCAGACTGATC 1645
Qy 1419 taagtgaatgaagtatccttaaaacaaagcatctattttggcagaaattgtgtct 1478
Db 1646 TAAATGTAATGTAAGTATCTTAAACAAAGCAATCTATTTTGGCAGAAATGTGTCT 1705
Qy 1479 taaattcagctatttgatctctgagacttcatatttctcatccctttattgtctttt 1538
Db 1706 TAAATTCAGTCATTGATATTCTGTGAGACTTCATATTCTCATCCCTTATTGCTTTT 1765
Qy 1539 agcaaacataagaaacacatgagtcattttgttcatttagatattctglaaaatctctg 1598
```

```
Db 1766 AGCAAAACATAAGAAACCATGAGTCATTTTGTCTATTAGAGTATTCTGATAAAATCTCTG 1825
Qy 1599 aaaaactactgaatacaaaaggttaagtattttttgttctatctctgatttgcattttattat 1658
Db 1826 AAAATACTGAATCAAAAGGTTAATGATTTTGTTCATCTGATTTGTCATTTATTATAT 1885
Qy 1659 ctgttatcggtgttaagtgtctaatttaccatttgatttttctgtctagacagataacttt 1718
Db 1886 CTGTTATCGGCTAAAGTGCTAAATTTACCCATTTTCATTTTCTGCTAGACAGATAACTTT 1945
Qy 1719 taatttttcaaatltygcagacacttttttttttttttttttttttttttttttttttttttt 1778
Db 1946 TAAATTTTCAAAATTTGGCAGACAC-TTTTFTTTTTTTTGAATAATCTTTCCCTTCAGATC 2004
Qy 1779 ttttgcccactgaacagcaccctccctcactgctccctggtgtccgattgggctggatgg 1838
Db 2005 TGTGCCCCACTGAACAGACCCCGTCCCTCACTGCTCTGGTGTCCGATTTGGCTCGATGG 2064
Qy 1839 tgttggggcatgatgtgtggagggaactggaaggtgcttttagtctggttcagggtcgggc 1898
Db 2065 TGTGGGGCATGATGTGTGGAGAACTGSAAGGTGCTTTAGGTCTGTGTTTCAGGGTCGGGC 2124
Qy 1899 attcttttctgttcgacacatctttttaaattttacaccttttcttaagaatctaatgcc 1958
Db 2125 ATCTCTTGTGTTGCACATCTTTTAAATTTTACACCTTTTCTTAAAGAAATCTTAATGCC 2184
Qy 1959 gcttaagtcttttataccaataatgctgagctttaaagtgtaggactctgtagtagacaca 2018
Db 2185 GCTTTAAGTTTTTATACCAATAATGCTGAGCTTTAAGTGTAGGATCTGTTAGTACAGACA 2244
Qy 2019 gtgtgatgga 2028
Db 2245 GTGTGATGGA 2254

RESULT 6
HSJ963E22/c
LOCUS
DEFINITION Human DNA sequence from clone RP5-963E22 on chromosome 20 Contains the 3' end of a novel gene similar to NY-REN-2 Antigen, 5 Cpg islands, ESTs, STSs and GSSs, complete sequence.
ACCESSION AL096828.29 GI:9663381
VERSION AL096828.29
KEYWORDS HTG: Cpg island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140152)
AUTHORS Lloyd,D.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Aug 2, 2000 this sequence version replaced gi:958454. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sv., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20 IMPORTANT: This sequence is not the entire insert of clone RP5-963E22 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
```


/note="7 copies 16 mer 66% conserved"
13006
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 500bp by BAMHI and ECORI digest data."
13276..13338
/note="3 copies 21 mer 85% conserved"
13844..13977
/note="MR58 repeat: matches 92. .233 of consensus"
16177..16269
/note="MLT1D repeat: matches 51. .146 of consensus"
16367..16629
/note="MLT1D repeat: matches 206. .494 of consensus"
17818..18094
/note="MLT1A1 repeat: matches 2. .278 of consensus"
18125..19582
/note="L1M4 repeat: matches 3996. .5549 of consensus"
19583..19986
/note="MSTA repeat: matches 1. .426 of consensus"

Query Match 42.9%; Score 870; DB 9; Length 140152;
Best Local Similarity 99.9%; Pred. No. 1.9e-190;
Matches 881; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1147 agaacggcagatcgaaacaaacatgaggcgaaacagtttcttcatgttttaacgt 1206
Db 106290 AGGAACGGCAGCTCGAACAACAAATGAGGGCGAACAGTTCTTACATGTTCTTAACGT 106231

QY 1207 ttgaactttgaaacagtttaaaacacacgtgcttggtgcagctccagtgctgcctcg 1266
Db 106230 TTGACTTTGAAACAGTTTAAACACGTGCTGTGGTCAGCTCCAGTGTCTGCTCCGCTG 106171

QY 1267 cgggggtgagtgctgcatcttgcccttctctgctgtgatttgcagatgagtcgt 1326
Db 106170 CGGGGTTGAGTGTGATCTTTGCCCTTCTCTGCTGTGATTTTGGCCAGATGATCTG 106111

QY 1327 cattatttctacttttctatgtattataatcctgttagaagtcactaaataaaggagtat 1386
Db 106110 CATTTATTCTACTTTTCTATGTATTAATCCTGTAGAAGTCACCTAATAAGGAGTA- 106050

QY 1387 tttttttgcagcttatcaatcagactgacttaagtgaagttaagtatccttaaaaa 1446
Db 106051 TTTTTTTGTCAGCTTATCAATCAATAGATGATCTAATGTGAATGAATGATCTCTTAAAAA 105992

QY 1447 caaagcatctattttggcagaaaattgttcttaattcagtcatttgattctgtgag 1506
Db 105991 CAAAGCATCTATTTTGGCAGAAATTGTCTTAAATTCAGTCAATTTGATATCTGTGAG 105932

QY 1507 acttcatttctcacccttattgtcttttttagcaacataagaacaccatgagtcattt 1566
Db 105931 ACTTCATATTTCTCATCCCTTTATTTGCTTTTAGCAACATAGAAACCATGAGTCATTT 105872

QY 1567 tgcatttagagtatctgataaaatccttggaaataactgaaatcaaaaagggttaagtat 1626
Db 105871 TGTCAATTTAGAGTATCTGTATAAATCTCTTGAAATACGTGAATCAAAAGGTTAATGAT 105812

QY 1627 tttttgtcattctgattgttcatttattatctgtctcgttcgaagtcactaattac 1686
Db 105811 TTTTGTGTCATCTGATTGTGCTATTTATCTGTTATCTGGTCTAAGTCTAATTTAC 105752

QY 1687 ccatttgattttctcgtagacagataaacttttaattttcacaatttggcagacactttt 1746
Db 105751 CCATTTGATTTTCTGCTAGACAGATAACTTTTAATTTTCAAAATTTGGCAGACACTTTT 105692

QY 1747 tttttttttgaaactcttctccagatctgttcccaactgaacagccaccctgcc 1806
Db 105691 TTTTTTTTTTGAAATCTTTTCTTCAGATCTGTGTGCCACTGAACAGCCACCCTGCC 105632

QY 1807 tcactgtcctggtgtccgattgggctggatggtgttggggcatgatgtgtggaggaaactg 1866
Db 105631 TCACGTGCTGCTGCTCGATTGGCTGGATGGTGTGGGGCATGATGTGTGGAGAACTG 105572

QY 1867 gaagggtctttagatctcgtgttcagggttcgggcatctcttctgtgttgcacatcttttaa 1926
|||||

Db 105571 GAAGGTGCTTAGTCTGTTTCAGGGTCGGGCATTCTTTGTTTGCACATCTTTTAA 105512

QY 1927 attttacacctttcttaagaattcttaataatccgcttaagtttttatataccaataatgctg 1986
|||||
Db 105511 ATTTTACACCTTTTCTTAAGAAATTTCTAATGCCGCTTAAGTTTATACCAATAATGCTG 105452

QY 1987 agctttaagtagtagatctgtagtcagacagacagtgtagtga 2028
|||||
Db 105451 AGCTTAAAGTCTAGATCTGCTAGTACACAGACAGTGTGATGCA 105410

RESULT 7
AC025565 168012 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 5 clone RP11-770E15, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC025565
VERSION AC025565.3 GI:9960368
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 168012)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 1, 2000 this sequence version replaced gi:7235344.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0770E15
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160169 bases at least Q40
Consensus quality: 162474 bases at least Q30
Consensus quality: 164047 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 166512; sum-of-contigs
Quality coverage: 4.42 in Q20 bases; agarose-fp
Quality coverage: 4.39 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 8432: contig of 8432 bp in length
* 8433: gap of unknown length
* 18030: contig of 9498 bp in length
* 18130: gap of unknown length
* 18131: contig of 8844 bp in length
* 26975: gap of unknown length
* 27075: contig of 17414 bp in length
* 44488: gap of unknown length
* 44589: contig of 12060 bp in length
* 56649: gap of unknown length

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 192994)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barne, N., Beckerly, R., Boguslavskiy, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurt, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, J., Rile, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 19, 2001 this sequence version replaced gi:7230119.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3524
Center clone name: 16_E18

----- Summary Statistics
Sequencing vector: M13; M7815; 52% of reads
Sequencing vector: Plasmid; n/a; 48% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181440 bases at least Q40
Consensus quality: 187416 bases at least Q30
Consensus quality: 189491 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 190394; sum-of-contigs
Quality coverage: 11.2 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 160053: contig of 160053 bp in length
* 160054 160153: gap of 100 bp
* 160154 160945: contig of 792 bp in length
* 160946 161045: gap of 100 bp
* 161046 161833: contig of 788 bp in length
* 161834 161933: gap of 100 bp
* 161934 162733: contig of 800 bp in length
* 162734 162833: gap of 100 bp
* 162834 163725: contig of 892 bp in length
* 163726 163825: gap of 100 bp
* 163826 164571: contig of 746 bp in length
* 164572 164671: gap of 100 bp
* 164672 165784: contig of 1113 bp in length
* 165785 165884: gap of 100 bp
* 165885 167095: contig of 1211 bp in length
* 167096 167195: gap of 100 bp
* 167196 167947: contig of 752 bp in length
* 167948 168047: gap of 100 bp
* 168048 169449: contig of 1402 bp in length
* 169450 169549: gap of 100 bp
* 169550 170722: contig of 1173 bp in length
* 170723 170822: gap of 100 bp

* 170823 171557: contig of 735 bp in length
* 171558 171657: gap of 100 bp
* 171658 172482: contig of 825 bp in length
* 172483 172582: gap of 100 bp
* 172583 173952: contig of 1370 bp in length
* 173953 174052: gap of 100 bp
* 174053 175031: contig of 979 bp in length
* 175032 175131: gap of 100 bp
* 175132 176068: contig of 937 bp in length
* 176069 176168: gap of 100 bp
* 176169 177688: contig of 1520 bp in length
* 177689 177788: gap of 100 bp
* 177789 179065: contig of 1277 bp in length
* 179066 179165: gap of 100 bp
* 179166 180432: contig of 1267 bp in length
* 180433 180532: gap of 100 bp
* 180533 181730: contig of 1198 bp in length
* 181731 181830: gap of 100 bp
* 181831 183033: contig of 1203 bp in length
* 183034 183133: gap of 100 bp
* 183134 184409: contig of 1276 bp in length
* 184410 184509: gap of 100 bp
* 184510 186294: contig of 1785 bp in length
* 186295 186394: gap of 100 bp
* 186395 188531: contig of 2137 bp in length
* 188532 188631: gap of 100 bp
* 188632 190451: contig of 1820 bp in length
* 190452 190551: gap of 100 bp
* 190552 192545: contig of 1994 bp in length
* 192546 192645: gap of 100 bp
* 192646 192994: contig of 349 bp in length.
FEATURES
Location/Qualifiers
source
1..192994
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-16E18"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1..160053
/note="assembly_fragment"
clone_end:SP6
vector_side:left
160154..160945
/note="assembly_fragment"
161046..161833
/note="assembly_fragment"
161934..162733
/note="assembly_fragment"
162834..163725
/note="assembly_fragment"
163826..164571
/note="assembly_fragment"
164672..165784
/note="assembly_fragment"
165885..167095
/note="assembly_fragment"
167196..167947
/note="assembly_fragment"
168048..169449
/note="assembly_fragment"
169550..170722
/note="assembly_fragment"
170823..171557
/note="assembly_fragment"
171658..172482
/note="assembly_fragment"
172583..173952
/note="assembly_fragment"
174053..175031
/note="assembly_fragment"
175132..176068
/note="assembly_fragment"
176169..177688
/note="assembly_fragment"


```
misc_feature 177789..179065
/note="assembly_fragment"
misc_feature 179166..180432
/note="assembly_fragment"
misc_feature 180533..181730
/note="assembly_fragment"
misc_feature 181831..183033
/note="assembly_fragment"
misc_feature 183134..184409
/note="assembly_fragment"
misc_feature 184510..186294
/note="assembly_fragment"
misc_feature 186395..188531
/note="assembly_fragment"
misc_feature 188632..190451
/note="assembly_fragment"
misc_feature 190552..192545
/note="assembly_fragment"
misc_feature 192646..192994
/note="assembly_fragment"
vector_side:right"
clone_end:T7
BASE COUNT 59747 a 38890 c 36686 g 55068 t 2603 others
ORIGIN

Query Match      18.0%  Score 365.8;  DB 2;  Length 192994;
Best Local Similarity 69.2%  Pred. No. 1.2e-73;
Matches 529;  Conservative 0;  Mismatches 232;  Indels 4;  Gaps 2;

QY 393  acccagagctcagctctccccagcacagccccagcttggctcaacgcgtagtaccag 452
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73613  ACCAACACACAGCAGCAACAAAGACCTCAGCCACAGCCCGCCTCACCAAGTGCAGCC 73554

QY 453  ccctcagcagcaccacccagaccctcgtgggtgcccacgacgaacagcggcgtttgg 512
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73553  TCAACAGCAGCAGCTGCAGATCCCTGGGTAGCTCCTCGTAAACAGGGGAGCAGCTTCAA 73494

QY 513  gcagagcggagggctggcagcagatagcaactctcctggaaacctcagccctaattctgc 572
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73493  CCAGAACATGAGCGGCGCAGTGAACACTTTGGTTAGGTGTTGACTCTGTCAGTGGCTTC 73434

QY 573  cccc---agegtcgaaatccccaccctccttgaataactgaagctgctcacagctcaca 629
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73433  ACCTCTTAGTGTAGAAGTGCATCCCGTCTGCGAAAGACTAAGGCCATTAACCAACTATA 73374

QY 630  cccgaagaggttgagtggaatctgaaagcggcgtgtgttctcatcatcaagagactactc 689
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73373  TCCCAAGAGCTTGTATTGGAATCTGAAGATGAGCGTGTGTTTATAATTAAAGCTACTC 73314

QY 690  tgagcagacatccaccgctccattagtagtactccatctggtgtgtagcagagcagcaca 749
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73313  TGAGGATGACATACATCGTTCCTTAATAACTCTATCTGGTGTAGTACTGACCATGTGTA 73254

QY 750  caagcgcctgacagcgccttccctgctcgtcgtcagcagcagggcgcgcctcactcgtctt 809
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73253  TAAGCGTTTGGATGCAGCTTACCCTGCTCAATGGGAAAGGCCCACTCTATTACTCTT 73194

QY 810  cagcgtcaaatggagtgaggcaatttttgggggtggcgcagatgaagtcctcccgtagacta 869
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73193  CAGTGTGAATGGCAGTGCACATTTTGTGGAGTGGCTCAAAATGAAGTCTGTGTGGAGCTA 73134

QY 870  cggcaccagtgccgggtcctggctcagcagcaagtgaaggggaagtttgatgtccagtg 929
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73133  TAATGCGGTATGCTGGTGTCTCAGGATAAGTGGGAGGGCAAAATTTGAAGTTAAATG 73074

QY 930  gattttgttaagatgtaccaccaataaccagctccgcgcacatccagcgtggagatacaga 989
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73073  GATCTTGTCAAGATGCTTCCCAATAACCAATTACGGCATATTCTGTTAGAAAATAATGA 73014

QY 990  caacaaacccggctcaacaaactcccggggacaccacccagagaggtgccttagaaaaagccaagca 1049
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73013  CAACAAACCGGTTACCAATTTCAAGGACACTCAAGAGAGTACCCCTTAGAAAAAGCTAAGCA 72954
```

```
QY 1050  agtctgaaattatcagttctctacaaagcacacacacctctcatcttcgacacttgcaca 1109
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 72953  AGTGCTTAAATAATATGCTACTTTCACAGCATACCACTCAATCTTTGATGACTTTGCACA 72894

QY 1110  ctacgaaagcgcg-ccagagagggagggtgggtgcgcgaaggaacgc 1153
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 72893  TTATGAAAGCGTCAAGAAGAGGAGGAGGAGCCATCGTAGGGTAAG 72849

RESULT 10
BC002559 2309 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, high-glucose-regulated protein 8, clone MGC:739
DEFINITION IMAGE:3139250, mRNA, complete cds.
ACCESSION BC002559
VERSION BC002559.1 GI:12803468
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2309)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: qcqpbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@hgrl.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tlonsong, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 4 Row: 1 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7705410.
FEATURES
Location/Qualifiers
source 1..2309
/organism="Homo sapiens"
/db_xref="LocusID:51441"
/db_xref="taxon:9606"
/clone="MGC:739 IMAGE:3139250"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
372..2111
/codon_start=1
/product="high-glucose-regulated protein 8"
/protein_id="AAH02559.1"
/db_xref="GI:12803468"
/translation="MSASSLLLEORPKGGKNGKNGVSHOKDGLNDDEPVLSPQARP
NNAYTAMSDSYLPSYSPGFSYSLGAEANSTGGDTAMPYLTYSYGLSNFPEFLPD
AMEQGPALGTPFLGQHGNNFFPSGIDFSAMGNSGOGSTSGSYSSNAYAPSSL
GGAMIDGQSAPANETLNKCMNTIDQMAALKIGSTEASNPKVGVSGSITS
NIVASNSLPPATIAAPPASWADIASKPAKQPKLTKNGTAGSSLPPPIKHNNDIG
```

TWNNKPVAKAPSAQALVONIGOPTGSPVVGQANNPPVAQASVGGTOLPLPPPPP
QPAOLSVQQAQPTTRWVAPRNRSGFHNGVGSQASGSGSTPSEHPVLLEKL
RSINNNPDKDNLKHGRVFIKISYSEDDIHRSIKYNIMSTGHRKRLDAAYRSMN
GKGPVLLFSYNGSCHFGVAPKMSAVDYNPTCAGVWSODKWKGRDFDVRWIFVKDVPNS
OLRHLRLNENKPVNTSRDTEQVPLEKAKQVLTIASYKHTTSIFDPSHYEKQEE
EESVKEHQGRGK"

BASE COUNT 613 a 581 c 595 g 520 t
ORIGIN
Query Match 16.8%; Score 341.6; DB 9; Length 2309;
Best Local Similarity 67.5%; Pred. No. 2.1e-68;
Matches 512; Conservative 0; Mismatches 239; Indels 7; Gaps 2;

Qy 416 cacagccccagcttggctcaacgcgcagatcatcagagccctcagcagccacccccagacc 475
Db 1360 CTCACCTCCACACAGCCTGCCAGCTTTCAGTCCAGCAACAGGCGAGCTCAGCCAACCC 1419
Qy 476 gctgggttccccacgcacagaaacgcggcgcttggcgagagcgaggggctggcagcg 535
Db 1420 GCTGGGTAGCACCTCGGAACCGTGGCAGTGGTTCGGTTCATATGCGGTGATGTAATG 1479
Qy 536 atagcaactctctggaaagcctcagcctaattctgccccccagcgtgaatcccccccg 595
Db 1480 GAGTAGGACAGCTCAGGCTGGTTCGGATCTACTCTCTTC-----AGAACCCCAACCCAG 1533
Qy 596 tccttgaaaaactgaaggtgctcacagctcaaacccgaaagagtttgagtggaactcga 655
Db 1534 TGTTCGAGAAAGCTTCGGTCCATTAATACTATAACCCCAAGATTTTGACTGGAATCTGA 1593
Qy 656 aaagcggcgctggttcatcatcaagagctactctgagagcacatccaccgctcccat 715
Db 1594 AACATGGCGCGGTTTTCATCATTAAGAGCTACTCTGAGGACGATATTCACCGTTCATT 1653
Qy 716 agtactccatctggtgtagcagagcagcgaacaaagcgctggacagcgcttcgcgt 775
Db 1654 AGTATAATATTTGGTGCACAGAGCATGGTAACAAGAGACTGGATGCTGTATCTGTT 1713
Qy 776 gcatgagcagcaggggcgctctaccctgctcagcgctcaatggagtgggcattttt 835
Db 1714 CCATGAACGGGAAGAGCCCGTTTACTTACTTTTCAGTGTCAACGGCAGTGACACTTCT 1773
Qy 836 gtggggtgcccagagatgaagtcctcccgctggactaagcgccacagtgccgggtctggtctc 895
Db 1774 GTGGCGTGCAGAAATGAATCTGCTGTGGACTTACAACACATCTGCAGGTGTGTGTCCTC 1833
Qy 896 aggacaagtggagaggaagtgttgcagtgagtgatgttttttaagagatgaccaca 955
Db 1834 AGACAAATGGGAAGGTCGCTTTTGATGTCAAGTGGATTTTGTGAAGGACGTTCCCAATA 1893
Qy 956 accagctccggcacatcagcgtggagaaataacgcacacaaacccggtcacaaactcccgg 1015
Db 1894 GCCAACTGCGACACATTCGCCCTAGAGAACACAGAGATAAACAGTACCACTCTAGGG 1953
Qy 1016 acacccagagtgcccttagaaaagcccaagcgaagtgtctgaaaaattatcagttcctaca 1075
Db 1954 ACACACTCAGGAAGTGCCTCTGGAAGAGGCTTAAGCAGGTGTGAAAATATATAGCCAGCTACA 2013
Qy 1076 agcacacaactccatctctcagcagcttgcctcactcagcagagcgcc-agagaggaag 1134
Db 2014 AGCACACCACTTCCATTTTGTATGATCTTCTCACACTATGAGAAACCCCAAGAGGAAG 2073
Qy 1135 aggtggtgcgaaggaacgcagagtcgaacaaacaa 1172
Db 2074 AAGTGTAAAGGACGCTCAAGGTCGTGGGAATAA 2111

RESULT 11
AX048109
LOCUS AX048109 2750 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 103 from Patent WO0070047.
ACCESSION AX048109
VERSION AX048109.1 GI:11876932

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Yang, H., Tang, Y.T., Lal, P., Reddy, R., Batra, S., Baughn, M.R.,
JOURNAL Yang, J., Azimzai, Y., Lu, D.A., Au-Young, J., and Shih, L.L.
Full-length molecules expressed in human tissues
Patent: WO 0070047-A 103 23-NOV-2000;
FEATURES Incyte Genomics, Inc. (US)
source Location/Qualifiers
1..2750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 1436854CB1"

BASE COUNT 776 a 619 c 638 g 717 t
ORIGIN
Query Match 16.8%; Score 341.6; DB 6; Length 2750;
Best Local Similarity 67.5%; Pred. No. 2.2e-68;
Matches 512; Conservative 0; Mismatches 239; Indels 7; Gaps 2;

Qy 416 cacagccccagcttggctcaacgcgcagatcatcagagccctcagcagccacccccagacc 475
Db 1162 CTCACCTCCACACAGCCTGCCAGCTTTCAGTCCAGCAACAGGCGAGCTCAGCCAACCC 1221
Qy 476 gctgggttccccacgcacagaaacgcggcgcttggcgagagcgaggggctggcagcg 535
Db 1222 GCTGGGTAGCACCTCGGAACCGTGGCAGTGGTTCGGTTCATATGCGGTGATGTAATG 1281
Qy 536 atagcaactctctggaaagcctcagcctaattctgccccccagcgtcgaaatcccccccg 595
Db 1282 GAGTAGGACAGCTCTCAGGCTGGTTCGGATCTACTCTTC-----AGAACCCCAACCCAG 1335
Qy 596 tccttgaaaaactgaaggtgctcagagctcaaacccgaaagagtttgagtggaactcga 655
Db 1336 TGTTCGAGAAAGCTTCGGTCCATTAATACTATAACCCCAAGATTTTGACTGGAATCTGA 1395
Qy 656 aaagcggcgctggttcatcatcaagagctactctgagagcacatccaccgctccat 715
Db 1396 AACATGGCGCGGTTTTCATCATTAAGAGCTACTCTGAGGACGATATTCACCGTTCATT 1455
Qy 716 agtactccatctggtgtagcagagcagcggaacaaagcgctggacagcgcttcgcgt 775
Db 1456 AGTATAATATTTGGTGCACGACAGCATGGTAACAAGAGACTGGATGCTGTATCTGTT 1515
Qy 776 gcatgagcagcaggggcgctctaccctgctctcagcgtcctaagtgaggagtgaccaca 835
Db 1516 CCATGAACGGGAAGAGCCCGTTTACTTACTTTTCAGTGTCAACGGCAGTGGACACTTCT 1575
Qy 836 gtggggtgcccagagatgaagtcctcccgctgagctacgcagccagtgccggggtcgtgctc 895
Db 1576 GTGGCGTGCAGAAATGAATCTGCTGTGGACTTACAAACACATGTGCAGGTGTGTGTCCTC 1635
Qy 896 aggacaagtggaggggaagtttgatgctcagtgaggtttttgttaaggagtgaccaca 955
Db 1636 AGGACAAATGGGAAGGTCGTTTGTGATGTCAGGTGGATTTTGTCAAGGAGCTGCCAATA 1695
Qy 956 accagctccggcacatcagcgtggagaaataacgcacacaaacccggtcacaaactcccgg 1015
Db 1696 GCCAACTGCGACACATTCGCCCTAGAGAACACAGAGAATAAACCCAGTGACCAACTCTAGGG 1755
Qy 1016 acacccagaggtgccccttagaaaaagcccaagcgaagtgtgaaaaattatcagttcctaca 1075
Db 1756 ACACCTCAGGAAGTGCCTCTGGAAGAGGCTAAGCAGGTGTGAAAATATATAGCCAGCTACA 1815
Qy 1076 agcacacaactccatctctcagcagcttgcctcactcagcagagcgcc-agagaggaag 1134
Db 1816 AGCACACCACTTCCATTTTGTATGATCTTCTCACACTATGAGAAACCCCAAGAGGAAG 1875
Qy 1135 aggtggtgcgaaggaacgcagagtcgaacaaacaa 1172

[illegible]

Best Local Similarity 74.2%; Pred. No. 4.4e-68;
Matches 443; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

QY 583 gaatccaccacccgctccttgaataactgaaggtgctcacagctacaacccgaagaagattt 642
Db 1286 GAGCCTCACCGGTTGGGAAACTTCGGTCCATTAATCACTAACCTAAAGATTTC 1345

QY 643 gagtgggaattgaataacgagcggtgtgtctcatcatcaagagctactctgaggagacatc 702
Db 1346 GACTGGAATCTGAACATGCGCGGGTTTCATCATTAAGAGCTACTCTGAGGACGATATC 1405

QY 703 caccgctccatttaagtactccatctggtgtagcagacagcagcgaacgagcgcctggac 762
Db 1406 CACCGTTCCTTAAGTATATCTGTGTGCGACAGAGCATGTTAACAAGACACTGGAT 1465

QY 763 agccctctcgcgtcagcagcaagggccggtctacctgctcttcagctcaatggg 822
Db 1466 GCCGCCTATCGTTCATGAATGGGAAGGTCCTGCTGACTTTCATGCTGCAACGGC 1525

QY 823 agtgggcatctttgtgggtggcgagatgaagtcctcccggtggactacggcaccagtgcc 882
Db 1526 AGTGGACACTTCTGTGAGTTGACAGATGAATCTGCTGGACTACAACACATGTGCA 1585

QY 883 ggggtctggtctcaggaacagtggaaggaagtttgatgtccagtgagttttgttaag 942
Db 1586 GGTGTGTGGTCCCAAGCAAAATGGAAGGGTGTTCGATGTTCAGATGGATTTTGTGAAG 1645

QY 943 gatgaccataaccacgtccggcacatcaggtgtgagataacgacacacaaacggctc 1002
Db 1646 GACGTTCCCAATAGCCAACTCCGACATCTGCTAGAGCAACAGGAGATAAACCAAGT 1705

QY 1003 acaactcccggggacacccagggaggtgctcttagaataagccaaagtgctgtaaaatt 1062
Db 1706 ACCAACTCTAGGACACTCAGGAAGTGCCCTGTGAAAAGCTAAGCAGGTTCGAATC 1765

QY 1063 atcagttctctaacgacacacacctccatctcttcagcagcttctcactacgagaagcgc 1122
Db 1766 ATACCCAGCTACAGACACACACCTTCATTTTGTGATGACTCTCACACTATGAGAAACGC 1825

QY 1123 c-agagagagagaggtggtgcgcaagaaagcgagagctgaaacaaacaaatgagg 1178
Db 1826 CAAGAGGAAGAGAAGTGTATAAAGAGACGTCAGGTCGTGGGAATAGAAAGCG 1882

RESULT 15
AC022766/c
LOCUS
DEFINITION Homo sapiens clone RP11-318E23, WORKING DRAFT SEQUENCE, 31
unordered pieces.
ACCESSION AC022766
VERSION AC022766.3 GI:8138013
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130984)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-318E23
JOURNAL Unpublished
2 (bases 1 to 130984)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bida,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vasiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2000 this sequence version replaced gi:7712161.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5852
Center clone name: 318_E_23
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 118950 bases at least Q40
Consensus quality: 124428 bases at least Q30
Consensus quality: 126573 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 127984; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1066: contig of 1066 bp in length
* 1067 1166: gap of 100 bp
* 1167 2747: contig of 1581 bp in length
* 2748 2847: gap of 100 bp
* 2848 4584: contig of 1737 bp in length
* 4585 4684: gap of 100 bp
* 4685 6485: contig of 1805 bp in length
* 6490 6589: gap of 100 bp
* 6590 8380: contig of 1791 bp in length
* 8381 8480: gap of 100 bp
* 8481 10218: contig of 1738 bp in length
* 10219 10318: gap of 100 bp
* 10319 12338: contig of 2020 bp in length
* 12339 12438: gap of 100 bp
* 12439 14278: contig of 1840 bp in length
* 14279 14378: gap of 100 bp
* 14379 16225: contig of 1847 bp in length
* 16226 16325: gap of 100 bp
* 16326 18627: contig of 2302 bp in length
* 18628 18727: gap of 100 bp
* 18728 21450: contig of 2723 bp in length
* 21451 21550: gap of 100 bp
* 21551 25430: contig of 3880 bp in length
* 25431 25530: gap of 100 bp
* 25531 28870: contig of 3340 bp in length
* 28871 28970: gap of 100 bp
* 28971 32940: contig of 3970 bp in length
* 32941 33040: gap of 100 bp
* 33041 35975: contig of 2935 bp in length
* 35976 36075: gap of 100 bp
* 36076 39460: contig of 3385 bp in length
* 39461 39560: gap of 100 bp
* 39561 43197: contig of 3637 bp in length

```
* 43198 43297: gap of 100 bp
* 43298 46333: contig of 3036 bp in length
* 46334 46433: gap of 100 bp
* 46434 51625: contig of 5192 bp in length
* 51626 51725: gap of 100 bp
* 51726 55173: contig of 3448 bp in length
* 55174 55273: gap of 100 bp
* 55274 59400: contig of 4127 bp in length
* 59401 59500: gap of 100 bp
* 59501 64807: contig of 5307 bp in length
* 64808 64907: gap of 100 bp
* 64908 68668: contig of 3761 bp in length
* 68669 68768: gap of 100 bp
* 68769 72524: contig of 3756 bp in length
* 72525 72624: gap of 100 bp
* 72625 77709: contig of 5085 bp in length
* 77710 77809: gap of 100 bp
* 77810 83970: contig of 6161 bp in length
* 83971 84070: gap of 100 bp
* 84071 89106: contig of 5036 bp in length
* 89107 89206: gap of 100 bp
* 89207 96647: contig of 7441 bp in length
* 96648 96747: gap of 100 bp
* 96748 105147: contig of 8400 bp in length
* 105148 105247: gap of 100 bp
* 105248 118840: contig of 13593 bp in length
* 118841 118940: gap of 100 bp
* 118941 130984: contig of 12044 bp in length.
```

FEATURES

source

```
1..130984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-318E23"
/clone_lib="RPC1-11 Human Male BAC"
1..1066
```

misc_feature

/note="assembly_fragment"

1167..2747

misc_feature

/note="assembly_fragment"

2848..4584

misc_feature

/note="assembly_fragment"

4685..6489

misc_feature

/note="assembly_fragment"

6590..8380

misc_feature

/note="assembly_fragment"

8481..10218

misc_feature

/note="assembly_fragment"

10319..12338

misc_feature

/note="assembly_fragment"

12439..14278

misc_feature

/note="assembly_fragment"

14379..16225

misc_feature

/note="assembly_fragment"

16326..18627

misc_feature

/note="assembly_fragment"

18748..21450

misc_feature

/note="assembly_fragment"

21551..25430

misc_feature

/note="assembly_fragment"

25531..28870

misc_feature

/note="assembly_fragment"

28971..32940

misc_feature

/note="assembly_fragment"

33041..35975

misc_feature

/note="assembly_fragment"

36076..39460

misc_feature

/note="assembly_fragment"

39561..43197

misc_feature

/note="assembly_fragment"

43298..46333

misc_feature

/note="assembly_fragment"

46434..51625

misc_feature

/note="assembly_fragment"

51746..55173

```
/note="assembly_fragment"
55274..59400
/note="assembly_fragment"
59501..64807
/note="assembly_fragment"
64908..68668
/note="assembly_fragment"
68769..72524
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
72625..77709
/note="assembly_fragment"
77810..83970
/note="assembly_fragment"
84071..89106
/note="assembly_fragment"
clone_end:T7
vector_side:right"
89207..96647
/note="assembly_fragment"
96748..105147
/note="assembly_fragment"
105248..118840
/note="assembly_fragment"
118941..130984
/note="assembly_fragment"
```

Query Match 16.4%; Score 333.4; DB 2; Length 130984;

Best Local Similarity 68.0%; Pred. No. 3.5e-66;

Matches 482; Conservative 0; Mismatches 221; Indels 6; Gaps 1;

```
QY 416 cacagccccagcttggctcaacgcagctatcagagccctcagcgcaccccccagccc 475
DB 82013 CTCACCTCCACACAGCCTGCCAGCTTTCAGTCCAGCAACGAGCAGCTACCCACCC 81954

QY 476 gctgggtgtcccccacgaacgaacgcgcgtttgggcagagcgaggggctggcagcg 535
DB 81953 GCTGGGTAGCAGCTCGGAACCGTGGCAGTGGTTCGTCATAATGGGTGGATGCTAA 81894

QY 536 atagcaactctcctggaaaagctccagcctaatcttgccccccagcgctcccccgcg 595
DB 81893 GAGTAGGACAGTCTCAGGCTGGTTCGATCTACTCTTC-----AGAACCCCA 81840

QY 596 tcttgaaaaactgaaggctgctcacagctacaacccgaaagagtttgagtgaaactga 655
DB 81839 TGTTCGAGAGAGCTTCGGTCCATTAACTAATAACCCCAAGAGTTTGTACTGAA 81780

QY 656 aaagcggcggtgtgttca tcaatgaagagctactctgagagcagacatccaccgcgt 715
DB 81779 AACATGGCGGGTTCATCATTAAGAGCTACTCTGAGGACGATATTCACCGTTCAT 81720

QY 716 agtactccatctggtgtagcagagcagcgcgaacaaagcgcctggacagcgcttcgc 775
DB 81719 AGTATAAATTTGGTGCAGCAGCAGCATGGTAACAAGAGACTGGATGCTGTAT 81660

QY 776 gcattgagcagcagggcccgctctacctgctcttcagcgtcaatggagtgaggcatttt 835
DB 81659 CCATGAACGGGAAGCCCGCTTACTTACTTTTCAGTGTCAACGSCAGTGGACACT 81600

QY 836 gtgggtgtccgagatgaagtcccccggtggaactagcagcagcagtgccggggtctgc 895
DB 81599 GTGGGTGGCAGAAATGAAATTCGTTGTGAGTGTGAGTGTGAGTGTGAGTGTG 81540

QY 896 aggacaagtggaaagggaagtttgatgtccagtgagtttttttaaggatgtaccacaa 955
DB 81539 AGGACAAATGGAAGGGTCCGTTTGTGATGTGAGTGTGAGTGTGAGTGTGAGT 81480

QY 956 accagctccggcacatcaggctggagaataacgacacaaacccggtcacacaaactcc 1015
DB 81479 GCCAACTGGACACATTCGCTTAGAGAACACAGAGAGAAACACGACGACACTCTAG 81420

QY 1016 acacccaggaggtgcctcttagaaaaagcccaagcgaagtgtgaaaaattatcagtt 1075
```

```

||||| ||||| ||||| | ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 81419 ACACTCAGGAAGTGCCTCTGGGAAAGGCTAAGCAGGTGTTGAAAATTATAGCCACCTACA 81360

Oy 1076 agcacacaacctccatcttcgacgactttgtcactacgagaagcgcca 1124
||||| || ||||| || ||||| . ||||| ||||| |||||
Db 81359 AGCACACCACTTCCATTTTGTGACTTCTCACACTATGAGAAAGGCCA 81311
```

Search completed: July 29, 2002, 18:09:38
Job time: 11004 sec

THIS PAGE BLANK (USPTO)